Investigating Knowledge and Use of Technical Vocabulary in Saudi Engineering Masters' Dissertations: A Corpus-based Study

Budur Alanazi

Thesis submitted to the University of Nottingham for the Degree of Doctor of Philosophy

October 2025

Abstract

This thesis explores technical vocabulary in engineering masters' dissertations, an academic genre that has received less attention compared to textbooks, course materials, and research articles. It addresses this gap through three linked studies, employing both quantitative and qualitative methods to investigate the knowledge and the usage of technical vocabulary in engineering masters' dissertations written by Saudi students. The first study examines the lexical profile and vocabulary load of a specialised corpus, consisting of 1,322,437 words compiled from such dissertations written by Saudi students at King Abdul-Aziz University. The findings showed that 4,000- and 7,000-word families, along with four supplementary lists and an engineering-specific word list, were necessary to achieve a minimal 95% and optimal 98% comprehension in engineering master's discourse. This highlights the lexically demanding nature of the Engineering Masters' Dissertation Corpus (EMDC).

The second study employs a combination of corpus-based and semantic approaches to identify and analyse single-word and multiword units (MWUs) of technical vocabulary in EMDC. The findings showed that 930 single-word items of engineering technical vocabulary (ETV) were identified, which covered 19.92% of tokens in the EMDC. The study found that 61.29% of the ETV (570 words) was from high-frequency vocabulary bands e.g., *bus, coolant, absorption*, 25.49% (237 words) from mid-frequency bands e.g., *aluminium, altitude, antennas*, and 4.12% (38 words) from low-frequency bands e.g., *brine, axial, photovoltaic*. Additionally, 2.0% (19 words) were from supplementary lists; proper nouns and transparent compounds, e.g., *Doppler, Gaussian, ciphertext*, and 7.1% (66 words) were from a specialised engineering word list e.g., *actuation, actuator, alloying*. The study also identified 856 MWU technical vocabulary items, which can be refined into 543 condensed MWUs e.g., *renewable energy, heat transfer, Nusselt number* from the EMDC. The distribution of ETV varied across dissertation sections: Results and Discussion sections contained the highest percentage (21.46%), followed by Methodology

(20.83%), while the Introduction (18.47%), Literature Review (18.44%), and Conclusion (18.56%) sections showed similar, lower percentages.

The third study investigates learners’ receptive knowledge of the ETV list and lecturers' evaluations of its usefulness. Participants were grouped into three proficiency levels–19 beginners, 33 intermediate, and 26 advanced learners–using the Updated Vocabulary Level Test (Webb, Sasao & Balance, 2017). Yes/No tests were administered to assess learners' receptive knowledge of ETV words selected from the 930 ETV list, while 20 engineering lecturers evaluated the usefulness of the ETV list. The study also analysed correlations between learners’ knowledge and teachers’ perceptions. The findings indicated that learners demonstrated stronger knowledge of high-frequency ETV compared to mid- and low-frequency words. Bootstrapped analysis showed that beginners, intermediate, and advanced learners knew progressively higher percentages of the target words, with advanced learners exhibiting the highest level of comprehension. Lecturers rated high-frequency ETV as the most useful, followed by mid- and low-frequency words. A strong correlation was observed between learners’ knowledge and lecturers’ perceptions, with the alignment improving as learners advanced in proficiency. This suggests that teachers’ perceptions become more accurate as learners develop greater vocabulary knowledge.

This thesis has both methodological and pedagogical implications for research on technical vocabulary in genre-specific written discourse and ESP. The developed wordlists can assist second language engineering students in learning the vocabulary essential for their field, enabling them to meet the language demands of their studies and professional practice. By providing targeted language support, students can enhance their proficiency and perform more effectively in their specialised areas.

Declaration

I declare that this thesis entitled [investigation knowledge and use of technical vocabulary in Saudi engineering masters' dissertations: a corpus-based study], is the result of my own work and investigation during my time as a PhD student at the university of Nottingham.

I confirm that the work presented compels with ethical guidelines of University of Nottingham, and all necessary ethical approvals have been obtained for the research conducted in May 2024.

Dedication

This thesis is dedicated to

my mother, Wadha, who passed away ten years ago after a brief battle with cancer.

May Allah grant her mercy and forgiveness. Amen.

Acknowledgements

The first word of thanks goes to Almighty Allah as with his blessings this journey with its joyful and difficult times was completed. I would like to express my sincerest gratitude to my two supervisors, Prof. Svenja Adolphs and Dr. Pawel Szudarski, for their invaluable guidance, great patience and continuous support during this academic journey. They took me from being a novice to an accomplished researcher. It has been an honour and a privilege to have grown with both of you.

I would also like to thank Northern Border University and the UK Saudi Arabian Cultural Bureau for my academic scholarship and the financial support to help me complete my PhD.

It is my honour to express my deepest gratitude to my loving family—my father, sisters, and brothers—for their unwavering love and encouragement throughout my life. Special thanks go to my two sisters, Tahany, who was by my side at the beginning of this journey, and Mahdiah, whose care for my baby boy was truly invaluable during the most demanding moments. Also, I am deeply grateful to my next-door neighbor Zainab's family for their kindness and support from the very first day I moved into my UK home and throughout my entire journey.

No words can truly express my deepest thanks to my beloved husband Dr. Saud, whose unwavering support and countless sacrifices throughout this journey made everything possible. His encouragement and belief in me never wavered. My heartfelt thanks also go to my dear children, Joud, Naif, Deema and Rayan; your pure souls and boundless pride in me have been my driving force. This achievement is as much yours as it is mine.

Table of Contents

Chapter 1 Introduction 1

1.1 Introduction 1

1.2 Study Context 1

1.2.1 Status of English in the Saudi Educational Context 2

1.2.2 English in Higher Education and Preparatory Programmes 3

1.2.3 English as Medium of Instruction in Engineering Programmes in Saudi Arabia 4

1.3 Overview of the Present the Thesis 7

1.4 Significance of Technical Vocabulary for Engineering Students 11

1.5 Conceptualization of Technical Vocabulary 12

1.5.1 Technical Multiword Units 13

1.6 Personal Interest, Motivation and Research Focus 15

1.7 Objectives of the Study and Research Questions 17

1.8 Structure of the Thesis 18

Chapter 2 : Literature Review 21

2.1 Introduction 21

2.2 Key Aspects of Vocabulary Knowledge 22

2.2.1 Receptive Vocabulary Knowledge 23

2.2.2 Technical Vocabulary Knowledge 23

2.3 Lexical Frequency Profiling 25

2.3.1 Vocabulary Load 26

2.4 Word Lists 28

2.4.1 Overview of Key Counting Units Used in Word List Development 29

2.4.2 Overview of Selection Criteria Adopted in Word List Generation 33

2.4.3 Types of Word Lists 37

2.5 A Disciplinary Approach to Written Discourse 60

2.5.1 Corpus-Based Lexical Analysis of Written Academic Genres 63

2.5.2 Dissertation/Thesis as a Distinctive Genre of Academic Writing....65

2.6 Teachers' Perceptions of Vocabulary Learning 68

2.7 Summary of the Chapter 69

Chapter 3 : Methodology 71

3.1 Introduction 71

3.2 Building the Engineering Masters' Dissertations Corpus (EMDC) 71

3.2.1 Constructing a Representative Corpus 71

3.2.2 Overview of the Engineering Masters' Dissertations Corpus 74

3.2.3 Corpus Cleaning 75

3.2.4 Piloting the Engineering Masters' Dissertations Corpus 77

3.3 Overview of Methodology in Study 1: Lexical Profiling of Vocabulary in the Engineering Masters' Dissertations Corpus 77

3.4 Overview of Methodology in Study 2: Identification of Technical Vocabulary in the EMDC 80

3.4.1 Identification of Single-Word-Unit Technical Vocabulary in the EMDC 81

3.4.2 Identification of Multiword Unit Technical Vocabulary Items in the EMDC 83

3.4.3 Analysis of Frequency Distribution of Technical Vocabulary Across the Sections of the EMDC 86

3.5 Overview of Methodology in Study 3: Teachers' Perception of Pedagogic Usefulness and Learners' Knowledge of Engineering Technical Vocabulary 87

3.5.1 Participants 89

3.5.2 Data Collection 89

3.5.3 Data Analysis 91

3.6 Ethical Issues 94

3.6.1 Informed Consent 95

3.6.2 Confidentiality and Anonymity 95

3.7 Summary of the Chapter 96

Chapter 4 : Study One: Lexical Profile and Vocabulary Load of the EMCD..........97

4.1 Introduction 97

4.2 The Target Corpus for Analysis 98

4.3 Analytic Approach: Lexical Profile Analysis..........99

4.3.1 Instruments for Analysis 101

4.3.2 Processing the Engineering Masters' Dissertations Corpus (EMDC) 102

4.4 Results and Discussion 108

4.4.1 Lexical Profile of the Engineering Masters' Dissertations Corpus (EMDC)..........109

4.4.2 Coverage of High-, Mid-, and Low-Frequency Vocabulary in the EMDC 113

4.4.3 Vocabulary Load of the Engineering Masters' Dissertations Corpus (EMDC) Across Nation's (2012) BNC/COCA Base Word Lists..........114

4.5 Summary of the Findings 120

Chapter 5 : Identification Engineering Technical Vocabulary..........123

5.1 Introduction 123

5.2 Identifying Single-Word-Unit Technical Vocabulary in the EMDC..........124

5.2.1 Keywords Analysis 125

5.2.2 Frequency Principles (Adopting the Frequency Cut-off Points)..........127

5.2.3 Consulting Engineering Dictionary and Concordance Lines..........128

5.2.4 Adapting the Semantic Rating Scale for Single-word Units..........128

5.2.5 Procedure for Analysing the Semantic Rating Scale 130

5.3 Identifying Technical Engineering Multiword Units in the EMDC..........132

5.3.1 Deciding the Length of Multiword Units..........133

5.3.2 Extracting the Engineering Multiword Units..........134

5.3.3 Determining Frequency Criterion for the Engineering Technical Multiword Units List 134

5.3.4 Determining Meaning Criterion for the Engineering Technical Multiword Units List..........135

5.3.5 Refining the Engineering Technical Multiword Units List..........137

5.4 Analysing of Technical Vocabulary Distribution in Different Sections of the Masters' Dissertations......138

5.5 Results and Discussion......139

5.5.1 Coverage of the Single-Word-Unit Technical Vocabulary in the EMDC......140

5.5.2 Distribution of Single-Word Unit Technical Vocabulary across High-, Mid-, and Low-Frequency Vocabulary Bands......144

5.5.3 Multiword Units Technical Vocabulary in the EMDC......148

5.5.4 Coverage of Single-Word Unit Technical Vocabulary Across Sections of Engineering Masters' Dissertations......156

5.6 Summary of the Findings and Rationale for the Next Chapter......160

Chapter 6 : Teachers' Perceptions of the Usefulness of Engineering Technical Vocabulary and Learners' receptive Knowledge......163

6.1 Introduction......163

6.1.1 Empirical Insights on Receptive Vocabulary Knowledge......164

6.2 Research Questions......166

6.3 Methodology......166

6.3.1 Participants......167

6.3.2 Data Collection......168

6.3.3 Analytic Approach......175

6.4 Results and Discussion......185

6.4.1 Saudi Engineering Students' Vocabulary Knowledge......186

6.4.2 Saudi Engineering Students' Receptive Knowledge of Engineering Technical Vocabulary......188

6.4.3 Receptive Knowledge of Engineering Technical Vocabulary Across Groups of Learners......194

6.4.4 Engineering Teachers' Evaluation of Pedagogic Usefulness of ETV List......196

6.4.5 The Relationship Between Learners' Vocabulary Knowledge and Teacher Perceptions Usefulness of ETV 204

6.5 Summary......206

Chapter 7 : Discussion and Conclusion......208

7.1 Introduction......208

7.2 The Lexical Profile Analysis of Engineering Masters' Dissertations Corpus (EMDC*)* **208**

7.3 Vocabulary load of the EMDC **211**

7.4 the Development of Single-word unit Engineering Technical Vocabulary Items **213**

7.5 Multiword Units Engineering Technical Vocabulary Items **216**

7.6 The Coverage of Engineering Technical Vocabulary Across the Five Main Sections of Masters' Dissertations **219**

7.7 Saudi undergraduate Students' Receptive Knowledge of Engineering Technical Vocabulary **222**

7.8 Teachers' Perceived Pedagogical Usefulness of Engineering Technical Vocabulary **224**

7.9 The Relationship Between Engineering Teachers' Perceived Usefulness of Technical Vocabulary and Students' Receptive Knowledge **227**

7.10 Contributions of the Study **229**

7.10.1 Methodological Contribution 229

7.10.2 Theoretical Contributions 232

7.11 Pedagogical Implications of the Study **234**

7.12 Limitations of the Study **237**

7.13 Concluding Remarks **239**

7.14 Recommendations for Future Research **240**

References **242**

Appendix A: **Teachers' Rating for Potential Engineering Technical Vocabulary List and Sematic Scale** **281**

Appendix B: **Teachers' Survey** **299**

Appendix C: **Sheet of Information for Participants** **304**

Appendix D: **Consent Form** **305**

Appendix E: **Condensed Engineering Technical Multiword Units** **306**

Appendix F: **Engineering Technical Vocabulary** **316**

Appendix G: **Engineering Technical Vocabulary Identified by the Raters.................322**

Appendix H: **YES/NO Receptive Vocabulary Tests 324**

Appendix I: 90 **Pseudowords Selected for Receptive Knowledge Test 330**

List of Tables

Table 2-1:Some of the Discipline-Specific Word Lists (Single Word Only)......................42

Table 2-2:Some of the Subject-Specific Word Lists (Single Word Only).........................44

Table 2-3: Technical/Discipline-Specific Multiword Items...48

Table 2-4: Engineering-Specific Word Lists 51

Table 3-1:Overview of the Cleaned Engineering Masters' Dissertations Corpus...........76

Table 3-2: EMDC Sub-corpora 86

Table 4-1: Nation's (2012) BNC/COCA Base Word Lists 101

Table 4-2: Bauer and Nation's Seven Levels of Word Family...103

Table 4-3: Lexical Profile of the EMDC across Nation's (2012) BNC/COCA Word Lists 109

Table 4-4: BNC/COCA Coverage of High-, Mid-, and Low-Frequency Vocabulary Across the EMDC 113

Table 4-5: Cumulative Lexical Coverage of the EMDC Across the BNC/COCA Base Word Lists to Reach 95% and 98% Lexical Thresholds 116

Table 5-1: A Modified Semantic Rating Scale Criterion 129

Table 5-2: The Most Frequent 20 Multiword Units 134

Table 5-3: Examples of Root Structures and the Variables...138

Table 5-4: The Coverage of Single-Word Unit Engineering Technical Vocabulary Across Nation's (2012) BNC/COCA Base Word Lists 141

Table 5-5: Distribution of Single-Word Unit Technical Vocabulary Across Schmitt and Schmitt's (2014) High-, Mid-, and Low-Frequency Vocabulary Bands 144

Table 5-6: Top 50 Single-Word Engineering Technical Vocabulary Across Nation's (2006) BNC/COCA base word lists 146

Table 5-7: Summary of the Condensed Multiword-Unit Technical Vocabulary Items in the EMDC...149

Table 5-8: The 20 Most Frequent Condensed Multiword Unit Technical Vocabulary Items in the EMDC...155

Table 5-9: Distribution of the Engineering Technical Vocabulary Across the Sections of Engineering Masters' Dissertations 157

Table 5-10: Distribution of Technical Vocabulary of High-, Mid-, and Low-Frequency in Each Section of Dissertations.....158

Table 6-1: Student Participants Information.....167

Table 6-2: Teachers' information.....168

Table 6-3: Distribution of Survey Items.....173

Table 6-4: Items Analysis of Yes/No Tests.....181

Table 6-5: Profile Analysis of the Items in Yes/No Tests.....182

Table 6-6: Learners' VLT Mean Scores (N = 78).....186

Table 6-7: Vocabulary Levels of Each Learner Group (N=78).....187

Table 6-8: Participants' Receptive Knowledge of ETV Based on the ISDT Scoring Method.....189

Table 6-9: Participants' Receptive Knowledge of ETV Based on Other Scoring Methods.....190

Table 6-10: Bootstrapped Descriptive Statistics of Students' Receptive Knowledge of Engineering Technical Vocabulary Based on Proficiency Levels.....194

Table 6-11: Teacher Evaluation of the Pedagogic Usefulness of Engineering TV list...197

Table 6-12: Top 20 Most Useful Words at Each Frequency Level.....200

Table 6-13: Correlation Between Learner Vocabulary Knowledge and Teacher Perception of Word Usefulness.....204

List of Figures

Figure 5-1: The concordance lines for the term 'Gunn diode' 136

Figure 5-2: Example of the concordance lines for the term 'research reactor' 153

Figure 6-1:Format of the VLT (Nation, 1983, 1990; Schmitt, Schmitt, & Clapham, 2001) .... **Error! Bookmark not defined.**

Figure 6-2: Yes/No test of the ETV list .... **Error! Bookmark not defined.**

Figure 6-3: Instructions for teacher surveys of word usefulness.....................................175

Figure 6-4: Three methods of correction for guessing in Huibregtse et al. (2002) 178

List of Appendices

Appendix A: Teachers' Rating for Potential Engineering Vocabulary and Semantic Rating Scale …………………………………………………………………………….280

Appendix B: Teachers' Survey ………………………………………………………..298

Appendix C: Sheet of Information for Participants…………………………………….303

Appendix D: Consent Form …………………………………………………………...304

Appendix E: Condensed Engineering Technical Multiword Units ……………………305

Appendix F: Engineering Technical Vocabulary ………………………………………315

Appendix G: Engineering Technical Vocabulary Identified by the Raters …………...321

Appendix H: Yes/ No Receptive Vocabulary Tests ……………………………………..323

Appendix I: Pseudowords Selected for Receptive Knowledge Test………………….....330

List of Abbreviations

EMDC	Engineering Masters' Dissertation Corpus
ETV	Engineering Technical Vocabulary
MWU	Multiword Unit
AVL	Academic Vocabulary List
AWL	Academic Word List
BASEWRD	Nation's (2012) BNC/COCA base word list
BNC	British National Corpus
COCA	Corpus of Contemporary American English
EAP	English for Academic Purposes
EFL	English as a foreign language
ESP	English for Specific Purposes
ESL	English as a Second Language
GSL	West's (1953) General Service List
L1	First language
L2	Second language
KAU	King Abdul-Aziz University
NBU	Northern Boarder University
KSA	Kingdom of Saudi Arabia
NAWL	New Academic Word List
VLT	Vocabulary Levels Test
UVLT	Updated Vocabulary Levels Test
ETV	Engineering Technical Vocabulary
AFL	Academic Formulas List

LFP	Lexical Frequency Profiling
CL	Corpus Linguistics
LLR	Log-Likelihood Ratio
I_{SDT}	Index of Signal Detection Theory
EMI	English Medium Instruction
KFUPM.	King Fahd University of Petroleum and Minerals
KSU	King Saud University
KAUST	King Abdullah University of Science and Technology
CIs	Confidence Intervals
EMI	English-Medium Instruction
PYP	Preparatory Year Programmes

Chapter 1 Introduction

1.1 Introduction

This chapter establishes the background and rationale for the present thesis, which aims to explore the usage of vocabulary in the engineering discipline from several perspectives. First, it deals with lexical coverage and vocabulary load from the specialised corpus of engineering masters' dissertations written by Saudi students. Second, identifying engineering technical vocabulary (ETV) in these dissertations adopts an 'innovative method' as proposed by Benson and Coxhead (2022) using corpus-based and semantic methods. Finally, it examines the receptive knowledge of the technical vocabulary of engineering students in an English as a foreign language (EFL) context. In addition, the thesis gauges lecturers' perceptions regarding the pedagogic usefulness of ETV. To this end, the present thesis is divided into three studies, which were conducted to address these aspects.

1.2 Study Context

This section provides the contextual background for the study by outlining the status of English in Saudi higher education and the implementation of English-Medium Instruction (EMI) in scientific programmes, with particular emphasis on its role in engineering education. It also highlights broader developments in Saudi educational policy and English language learning, situating the investigation of technical vocabulary use in masters' dissertations within the engineering discipline. Finally, this section conceptualizes the notion of technical vocabulary as it applies to engineering contexts, providing a foundation for the study's focus on its use in masters' dissertations. Understanding how such vocabulary is defined and operationalized is crucial for investigating its role in engineering education, where mastery of discipline-specific terminology directly influences students' academic success and professional readiness.

1.2.1 Status of English in the Saudi Educational Context

Since its establishment in 1932, Saudi Arabia has pursued rapid development in all fields, including education. The Kingdom's education policies, overseen by the Ministry of Education, have continually evolved, with significant attention given to English language education. English is taught as a foreign language in Saudi Arabia, aligning with Kachru's (1992) "expanding circle" framework, where the language is primarily learned for international communication and academic purposes rather than for internal use.

The importance of English in Saudi Arabia is widely recognized by the government, educators, and students alike. As a global lingua franca, English is essential for international business, industry, and academia, providing access to scientific publications and supporting cross-cultural interactions. This recognition has led to English being a compulsory subject in all Saudi schools since 2021.

Despite its compulsory status and significant government investment, the quality of English education in Saudi Arabia faces challenges. While many students are admitted to university-level English departments, many graduates still struggle to produce coherent spoken or written sentences (Alharbi, 2021). This has raised debates about teacher quality and instructional methods. To address these concerns, large numbers of expatriate teachers are recruited annually, and many Saudi students travel abroad to Western countries to strengthen their English skills (Tamran, 2016). The government, educators, and students increasingly view English not simply as a subject but as a vital tool for higher education, international communication, and professional success.

The transformation of English language education is also aligned with Saudi Vision 2030, launched in 2016, which aims to build a vibrant society, a strong economy, and an ambitious nation (Vision 2030, 2019). Vision 2030 explicitly recognizes English as a global language necessary for commerce, science, and technology, and calls for educational reforms that enhance students' proficiency. As Elyas and Badawood (2017) note, English programmes are supported through free education, student stipends, and system-wide policies designed by the Ministry of Education. The overarching goals of English teaching include equipping students with core language skills, preparing them for real-life communication, fostering positive attitudes toward English, and ensuring they can use the language effectively in professional fields (ur Rahman & Alhaisoni, 2013).

1.2.2 English in Higher Education and Preparatory Programmes

English assumes an even more critical role in Saudi Arabian higher education. Most Saudi public universities employ EMI in disciplines such as engineering, medicine, business administration, computer science, information sciences, and applied sciences (Al-fehaid, 2018). However, students entering these programmes typically graduate from government schools where English is taught only as a foreign language. Despite nine years of instruction (from grade 6 to grade 12), many lack the proficiency required for success in university-level courses (Al Asmari & malik, 2016). To address this gap, nearly all universities have introduced one-year Preparatory Year Programmes (PYPs) to develop students' English communication, study, and computer skills (Muhammad & Abdul Raof, 2020).

Universities across the Kingdom have introduced PYPs that include English, mathematics, and general sciences. These programmes are designed to equip students with the necessary language and scientific skills before they begin their main degree programmes. English is central to the PYPs, as most scientific disciplines, including medicine, engineering, and the natural sciences, are taught through EMI. This strategic shift underscores the critical need for students to achieve a high level of English proficiency, particularly in mastering the technical vocabulary essential to their engineering studies and future careers.

This study is situated within this context, investigating Saudi engineering students' knowledge and use of technical vocabulary in masters' dissertations, which represent a primary form of academic writing within this specialised genre. The research aims to provide insights that can help improve engineering programmes delivered through English in Saudi higher education, thereby supporting the goals of Vision 2030. For engineering students in particular, EMI provides access to international knowledge and technical resources but also introduces challenges, especially in mastering specialised vocabulary. Understanding these challenges requires closer examination of how English functions as the medium of instruction in Saudi higher education and how students engage with technical vocabulary in academic genres such as masters' dissertations. The next section, therefore, examines the role of EMI in Saudi universities, with specific reference to engineering education.

1.2.3 English as Medium of Instruction in Engineering Programmes in Saudi Arabia

EMI has played a pivotal role in Saudi Arabian higher education. As a global phenomenon, EMI has reshaped teaching and learning practices, particularly in non-English-speaking countries (Siegel, 2020). In Saudi Arabia, this transition is closely linked to national educational reforms aimed at economic diversification and global integration (Barnawi & Al-Hawsawi, 2017). The adoption of EMI reflects a deliberate strategy to align the nation's educational system with international standards while addressing local socio-economic and cultural priorities (Alshahrani, 2020).

The spread of EMI in Saudi universities, particularly in scientific fields, has drawn significant scholarly attention (De Jong, 2018). EMI refers to the use of English to teach academic subjects in contexts where it is not the majority's first language (Macaro, Tian & Chu, 2020). While EMI supports internationalization, it also presents challenges. Rapid implementation has strained the recruitment of qualified staff and exposed the limitations of students' English proficiency, with consequences for academic achievement and cultural identity (Louber & Troudi, 2019). Students often struggle with comprehension and communication, while faculty report difficulties teaching exclusively in English, leading many to rely on Arabic (Graham & Eslami, 2019). Gaffas (2019) found that many Saudi undergraduates were not prepared for EMI. English use was often confined to technical terminology, while Arabic dominated classroom discussion. As Graham and Eslami (2019) emphasize, more research is needed to understand EMI's global impact. Based on Kahu and Nelson's (2018) framework, institutional policy and curriculum design are also critical in shaping student engagement and success in EMI contexts.

In engineering education, EMI supports national and institutional initiatives to enhance technical training and ensure future engineers can engage in international research and innovation while adapting knowledge to local contexts. English proficiency is increasingly viewed as essential for professional advancement, reflecting wider cultural and professional shifts. This trend is visible in the mobility of engineering students, international collaboration in research and industry, and the rising demand for English skills in both domestic and global job markets (Macaro et al., 2020). While EMI in Saudi engineering programmes raises concerns about linguistic readiness and the status of Arabic, it underscores the need for further empirical research (Alanazi & Curle, 2025). Addressing this gap, the present study investigates the use of technical vocabulary in the masters' dissertations of Saudi engineering students, offering insights for shaping effective pedagogical and institutional strategies.

At the undergraduate level, Saudi universities such as King Fahd University of Petroleum and Minerals (KFUPM), King Saud University (KSU), King Abdulaziz University (KAU), and King Abdullah University of Science and Technology (KAUST) offer a wide range of engineering programmes through EMI. Core disciplines such as civil, mechanical, electrical, and chemical engineering are available across leading public universities, alongside specialised tracks including industrial and systems engineering, computer and software engineering, biomedical engineering, and environmental engineering. These programmes are designed to prepare students for the technical and professional demands of the global market, aligning with national development priorities and Vision 2030 initiatives. However, since most entrants come from government schools where English is taught as a foreign language, students often face linguistic challenges in engaging with EMI content from the outset of their studies.

English language proficiency remains a challenge for students at the tertiary level in Saudi Arabia (Al-Seghayer, 2021). Many students enroll in EMI engineering programmes with limited English skills, hindering their ability to engage with complex course material and reducing the programs' long-term effectiveness. Studies consistently link English proficiency with academic success in EMI contexts. At the undergraduate level, proficiency is essential for comprehending lectures, understanding technical textbooks, conducting laboratory work, and engaging in group projects (Alrabai, 2018). However, many of these skills are not adequately covered in the PYP. Engineering students, in particular, face challenges in understanding discipline-specific vocabulary embedded in textbooks written in English, which often exceeds their proficiency level.

At the postgraduate level, Saudi universities also offer master's and PhD programmes across these engineering fields, with coursework and research requirements delivered in English. As students advance to master's programmes, proficiency becomes even more critical for engaging with research articles, writing literature reviews, and producing technical reports—culminating in the dissertation (Alhamami, 2021). At the doctoral level, English dominates academic life, serving as the primary medium for designing and disseminating research, publishing in international journals, and participating in conferences—factors that directly influence research quality and international visibility (Alasmari & Khan, 2020).

Institutions such as KFUPM, KSU, KAU, and KAUST provide master's degrees in civil, mechanical, chemical, electrical, and computer engineering, in addition to specialised programmes in areas such as biomedical and environmental engineering. These master's programmes are research-intensive and require students to engage critically with technical literature, develop advanced academic writing skills, and contribute to scholarly and industrial knowledge through dissertations and, increasingly, publications in international journals.

The progression from undergraduate to postgraduate engineering programmes highlights the growing linguistic and academic challenges faced by Saudi learners. At the undergraduate level, many students struggle with low English proficiency and limited exposure to ESP, which hinders their ability to acquire and apply technical vocabulary effectively. These challenges intensify at the master's level, where students are expected not only to master discipline-specific terminology but also to demonstrate advanced academic literacy in producing dissertations and research articles. The linguistic gap between undergraduate preparation and postgraduate expectations often results in difficulties in both comprehending and producing technical texts, underscoring the critical need for targeted ESP support and vocabulary development throughout engineering education in EMI contexts.

1.3 Overview of the Present the Thesis

Vocabulary is the most essential of all linguistic units and the one which speakers of a language must know to communicate (Nation, 2006; Taylor, 2015). It is a crucial element of learning a language, especially for second/foreign language learners (Csomay& Petrovic 2012). Access to the knowledge of vocabulary is crucial for the attainment of fluency and competency in comprehension and language production (Schmitt, 2010). Although grammar is crucial for clear communication, without vocabulary, nothing can be communicated (Wilkins, 1972). Research on lexical profiles illustrates the importance of vocabulary in language learning by exploring the coverage of words in a text and how much coverage is needed by learners to understand a particular text (Szudarski, 2018).

The integration of corpus linguistics, along with the advancement of vocabulary analysis tools, such as RANGE, AntWordProfiler, and Vocabprofile, has noticeably contributed to research in the field of vocabulary studies. Corpus linguistics refers to the study of collections of "naturally occurring language texts chosen to represent a particular state or variety of language" (Sinclair, 1991, p. 171). In recent years, corpora have been widely utilized to examine extensive datasets, enabling detailed vocabulary profile analysis of texts and reducing reliance on traditional intuition-based methods (Cheng 2012; Nurmukhamedov & Webb, 2019). Employing corpus-based analysis to analyse a larger collection of data enables researchers to develop a list of word frequencies, and these have been used together with software to profile the distribution of words within texts at different frequency levels (Nurmukhamedov & Webb, 2019). Another important contribution of corpora to vocabulary studies is that they help provide information about text difficulty and determine the amount of vocabulary needed by learners at different levels to understand different types of spoken and written discourse (Nurmukhamedov & Webb, 2019). Several notable word lists derived from corpora have been published over the past two decades (Watson Todd, 2017). These lists are primarily designed for pedagogical purposes, such as helping teachers develop vocabulary learning objectives for learners in line with the notion that a word's frequency reflects its usefulness for learners (Coxhead, 2011). While they are important for setting vocabulary assessment tasks, analyzing text difficulty, and improving learning materials (Gardner & Davies, 2014), their practicality for selecting vocabulary for explicit instruction in a course is limited as most lists contain over 2,000 words. Although the current study is not an experimental investigation into teaching vocabulary, it might offer valuable insights that could help address the issues mentioned earlier. Specifically, this study aims to examine EFL students' receptive knowledge of technical vocabulary and teachers' perceptions of its relevance in engineering discourse. Additionally, it provides pedagogical perspectives on vocabulary knowledge and usage in the university context. A strong command of the technical vocabulary in a specific field and the ability to use it effectively demonstrate a profound understanding of the subject (Coxhead, McLaughlin & Reid, 2018). Therefore, ‘content knowledge’ and ‘specialised lexical knowledge’ are deeply intertwined (Woodward-Kron, 2008). Reid, Coxhead, and McLaughlin (2018) define technical vocabulary as terms typically found in areas of specialization, and most people outside the sector are unfamiliar with them (Nation 2013). For example, welders may be familiar with terms like *polyurethane*, *scalars*, and *rutile*, but these terms are not often used outside the trade. This means that students must concentrate on both the course material and the technical vocabulary at the same time (Coxhead & Demecheleer, 2018).

Liu and Lei (2020) point out that technical vocabulary deserves special focus in language learning because many technical words occur frequently in general language, and even highly common polysemic technical words often have specialised meanings that differ from their everyday usage. This causes significant challenges for L2 learners in comprehending such vocabulary (Coxhead, Demecheleer & McLaughlin, 2016). Another reason is that technical vocabulary items are often opaque in meaning, making them particularly difficult for EFL learners to understand (Watson Todd, 2017). Additionally, the rapid evolution of various fields and technologies leads to the constant introduction of new technical terms, underscoring the importance of dedicated attention to this type of vocabulary (Liu & Lei, 2020).

Another important aspect is the disciplinary and genre approach to vocabulary study. Thompson and Hunston (2020) highlight that the focus of research in English for academic purposes (EAP) has evolved, shifting from an initial emphasis on register analysis to a broader exploration of discourse and genre analysis. This includes examining patterns, rhetorical gestures, posture, and phraseology. This shift has led to an increased emphasis on the concept of 'disciplines', where subject areas are defined not only by specialised terminology and 'grammatical structures' but also by distinct discourse and cultural perspectives (Thompson & Hunston, 2020). EAP practitioners, particularly corpus linguists (e.g., Hyland, 2000; Thompson, 2016), have played a key role in advancing research on disciplinary discourses. For instance, Hyland (2000) has examined a range of text genres across disciplines such as soft and hard sciences and applied linguistics. The present study particularly focuses on dissertations as a sub-genre in academic contexts. Most previous studies were on research articles as a genre (Thompson, 2016; Hyland &, Shaw, 2016). Hyland and Shaw (2016) observed that theses and dissertations are categorized as research genres within academic discourse, though they are generally less interactive and multimodal compared to research articles and textbooks. Research students, especially at the postgraduate level, are required to produce texts such as research articles and conference papers, before completing their theses (Hyland & Shaw, 2016). Along the same lines, D'Angelo (2016) has pointed out that many students and early-career researchers face challenges when creating academic posters to present their preliminary research findings in their respective fields.

1.4 Significance of Technical Vocabulary for Engineering Students

The selection of vocabulary for instructional focus should be guided by two primary considerations: learners' needs and the usefulness of the vocabulary items (Schmitt & Rodgers, 2020). For engineering students, familiarity with technical vocabulary is essential. Technical vocabulary occurs frequently within specialised disciplines and is indispensable for students and professionals engaging with technical content (Liu & Lei, 2020). Such vocabulary is often unique to a specific field, such as engineering or medicine, and is rarely used in other academic or professional contexts (Chung & Nation, 2004; Nation, 2008, 2013; Coxhead, 2013). It serves to encapsulate discipline-specific concepts and is typically acquired by specialists within that field (Coxhead, 2018). Moreover, knowledge of technical vocabulary facilitates entry into specific discourse communities and is linked to scientific literacy (Gablasova, 2014).

In the Saudi context, engineering students face additional challenges in acquiring technical vocabulary due to linguistic and cultural differences, along with limited opportunities for experiential English exposure, particularly in EMI programmes (Al Roomy & Alhawsawi, 2019). Mastery of technical terms is crucial not only for comprehending engineering content but also for identifying and using new terminology within the discipline. Those who engage deeply with a field become part of a "system of subject knowledge" (Chung & Nation, 2003, p. 252), sharing both technical vocabulary and disciplinary norms that allow them to access and contribute to domain-specific discourse (Coxhead, 2013; Woodward-Kron, 2008). As Coxhead (2013, p. 116) notes, "People outside that academic or professional sphere might have some knowledge of this vocabulary but the people inside these areas of language use would be expected to be able to understand and use this language fluently."

For L2 engineering students in EMI settings, in addition to practicing engineers, awareness and accurate use of technical vocabulary are vital for acquiring subject knowledge and communicating effectively with peers. Proficiency in technical vocabulary enhances the efficiency and quality of professional communication. This study aims to provide insights that can help Saudi EFL learners improve their acquisition of technical vocabulary. Given that technical vocabulary accounts for a substantial portion of disciplinary texts, mastery of these terms supports reading comprehension and vocabulary development across diverse contexts (Sun & Dang, 2020). It is therefore essential for EFL learners in specialised fields like engineering to develop a strong foundation in technical vocabulary before progressing to lower-frequency words (Lu, 2018).

1.5 Conceptualization of Technical Vocabulary

The concept of technical vocabulary is referred to by various terms, including 'discipline-specific', 'specialised vocabulary' and 'semi-technical vocabulary', depending on the research focus or objectives (Coxhead et al., 2018). Technical vocabulary refers to subject-specific terms used to convey specialised knowledge in a particular field of study or domain (Liu & Lei, 2020; Omidian & Siyanova-Chanturia, 2021). Technical vocabulary refers to the specialised terminology used in a particular field of study or professional practice (Nesi, 2013; Szudarski, 2018). These domain-specific lexical items are used by members of a particular academic or professional community to create a specialised discourse tailored to their audiences, often inaccessible to individuals outside their discipline (Omidian & Siyanova-Chanturia, 2021). In other words, technical vocabulary refers to single-word or multiword units with meanings that are specialised to a particular discipline, such as engineering and sciences (Tongpoon-Patanasorn, 2018). According to Nation (2001), technical vocabulary refers to words that are closely associated with a specific topic, field, or domain. An alternative approach to defining technical vocabulary focuses on usage patterns, differentiating between general and field-specific contexts. From this perspective, Nation (2001) organizes words into frequency bands and identifies a word as technical if its core meaning is primarily used in a specialised field, even if it occasionally appears in general language contexts. The significance of technical vocabulary in specialised texts is substantial, with studies indicating it constitutes 20–30% of the content (Chung & Nation, 2003). This type of vocabulary, defined by its strong association with a specific field, is as essential for specialised learners as high-frequency words are for general learners. Therefore, an effective approach to teaching it must combine deliberate, focused attention with ample opportunities for learners to encounter and use the terms in context (Nation & Meara, 2002).

Considering the aim of the present study, it is important to clarify common misconceptions about technical vocabulary. As noted by Benson and Coxhead (2022), Chung and Nation (2003), Lu (2018), and Nation (2013), technical vocabulary is not restricted to rare or highly specialised terms. However, some studies, such as Hamied, Sundayana, and Kwary's (2019) study, focused on such words. Technical vocabulary can be words listed in the General Service List or the Academic Word List (Tongpoon-Patanasorn, 2018). In other words, technical vocabulary can be words listed in Schmitt and Schmitt's (2014) high-, mid-, and low-frequency lists (Benson, 2020; Benson & Coxhead, 2022; Chung & Nation, 2003; Nation, 2013; Nation, 2016). This point is important, because everyday words may also be technical, e.g., the use of *flow* in plumbing (Coxhead & Demecheleer, 2018). The two terms, specialised and technical vocabulary, are used interchangeably in this thesis.

1.5.1 Technical Multiword Units

The term multiword units (MWUs) has been used in the literature interchangeably with other related terminologies, such as formulaic language (Schmitt, 2010; Wray, 2000); formulaic expressions (Biber, Conrad & Cortes, 2004), ‘phraseological chunks’ or ‘academic phraseology’ (Szudarski, 2018), ‘n-grams’, ‘lexical bundles’, ‘multiword units’, lexical collocations (Siyanova-Chanturia & Omidian, 2020; Wood, 2020). These combinations are essential for understanding and producing language accurately in specialised fields. Nation (2016, p. 71) defines multiword items as “phrases that are made up of words that frequently occur together”. These include common two-word collocations e.g., *data analysis, clinical trial* (Shin & Nation, 2008), phrases of three or more words (Coxhead, 2018), and phrasal verbs or expressions (Garnier & Schmitt, 2015; Martinez & Schmitt, 2012).

The relationships between these words are not arbitrary; rather, they are governed by predictable patterns and statistical regularities. Extensive research has been conducted to explore and quantify these linguistic connections, highlighting how words naturally co-occur and form meaningful collocations (Sinclair, 1991). However, a researcher should decide which terminology is relevant to their research purpose. For specialised lists, Siyanova-Chanturia and Omidian (2020, p. 530) prefer to use multiword items/expressions/units instead of formulaic language. The term ‘multiword’ specifically refers to linguistic sequences that consist of more than one word (Wray, 2002). In contrast, ‘formulaic language’ encompasses both single-word and multiword expressions, including conversational speech formulas e.g., *yes, hello*, expletives and exclamations e.g., *darn, wow*, and other fixed expressions at the single-word level (Van Lancker-Sidtis & Rallon, 2004). Therefore, in the present study, the term multiword unit (MWU) is used to refer to the phrases containing two to five words found in the target corpus.

1.6 Personal Interest, Motivation and Research Focus

The motivations for this study were driven by observations regarding the researcher's MA dissertation which was conducted in fulfilment of the requirement for a master's degree in applied linguistics. In this dissertation, the researcher examined the usage of academic vocabulary by comparing EFL learners with native English speakers in their academic assignments. However, for the research for this PhD, the researcher was self-motivated to develop an interest in further investigating the use of technical vocabulary in Saudi students' masters' dissertations in the engineering discipline. This is also a result of her initial perceptions as an L2 lecturer of English, and her experience while studying in the UK, where she interacted with many other L1 and L2 English postgraduate students and observed that students faced a challenge in using vocabulary, especially those that carried specialised meanings in English.

Recent research on lexical variation has highlighted the importance of word lists in English for Specific Purposes (ESP), utilizing a range of corpus-based approaches, including lexical frequency profiling, keyword analysis, and concordance line analysis, to investigate technical vocabulary in various disciplines (e.g., Benson, 2020; Benson & Coxhead, 2022; Lu, 2018). The current thesis intends to explore the use of technical vocabulary in engineering masters' dissertations written by Saudi students, employing both corpus-based and semantic-based methodologies.

This thesis has three overarching aims. First, it intends to analyse the lexical profile of masters' dissertations in the field of engineering. This objective was addressed in Study 1 through lexical profiling and assessing the vocabulary load of a newly developed specialised corpus, the engineering masters' dissertation corpus (EMDC), as detailed in Chapter 3 (see Section

3.2). Second, it seeks to explore and identify ETV and their distribution within the EMDC. This was addressed in Study 2 using corpus-based and semantic-based approaches to analyse the vocabulary in the EMDC and identify ETV, including both 'single-word unit' and 'multiword unit' items. This led to the creation of pedagogically oriented lists of technical single-word and multiword units, intended for use in university engineering courses by both instructors and students. The single-word list of the ETV includes terms such as *colloids, atomic, beam, convective, modulus, tensil*e, and *composites.*

While the importance of technical single-words is well-documented (Nation, 2013), later studies have increasingly highlighted the significance of multiword units in language acquisition and usage (Benson, 2020; Benson & Coxhead, 2022; Lu, 2018; Biber, 2009; Hyland, 2008). However, studies specifically addressing technical multiword units in engineering remain limited. Notable exceptions include Ward's (2007) work on chemical engineering textbooks, Wood and Appel's (2014) analysis of first-year business and engineering textbooks, and Fox and Tigchelaar's (2015) research on published engineering articles. The present study intends to contribute to this growing body of research by examining ETV, including both single-word terms and multi-word units, found in masters' dissertations in the field of engineering.

Finally, this thesis aims to explore learners' receptive knowledge of the ETV items and lecturers' evaluation of their pedagogic usefulness. The knowledge of technical vocabulary is essential for learners in an ESP context for two reasons. First, mastering specialised vocabulary is essential for L2 learners to integrate into specific professional communities (Benson & Coxhead, 2022; Benson, 2020; Coxhead, 2013; Dang, et al., 2022a, b; Wray, 2002). Second, these vocabulary lists provide teachers with a practical tool to determine which words to prioritize, the rationale behind teaching them, and effective ways of incorporating the lists into language instruction (Nation, Coxhead, Chung & Quero, 2016). Investigating learners' knowledge of ETV and lecturers' evaluations of these words' usefulness is pivotal in understanding the technical nature of the engineering lexicon.

1.7 Objectives of the Study and Research Questions

The overarching aim of this thesis is to investigate the knowledge and use of technical vocabulary in masters' dissertations written by Saudi engineering students. For this study to achieve objectives outlined below, the following research questions will provide a framework:

1. What is the lexical profile of the EMDC across Nation's (2012) BNC/COCA word frequency lists?
    a. What is the coverage of Schmitt and Schmitt's (2014) high-, mid-, and low-frequency vocabulary distributed in the EMDC?
2. What is the vocabulary load of EMDC across Nation's (2012) BNC/COCA word frequency lists?
3. Which single-word unit technical vocabulary items from the EMDC are used in Saudi EFL students' writing?
    a. How many of these items fall into the high-, mid-, and low-frequency vocabulary bands, respectively?
4. Which multiword unit technical vocabulary items from the EMDC are used in Saudi EFL students' writing?
5. How is the ETV list distributed across the five different sections (Introduction, Literature Review, Methodology, Results and Discussion, and Conclusion) of an engineering masters' dissertation?
6. To what extent do Saudi undergraduate engineering students understand the technical vocabulary in their field of study receptively?

7. Which technical engineering vocabulary items do teachers perceive as being useful for pedagogical purposes?
8. Is there a relationship between engineering teachers' perceptions of the usefulness of the ETV and the receptive knowledge of this vocabulary among Saudi undergraduate engineering students?

1.8 Structure of the Thesis

The present research comprises seven chapters. Chapter 1 provides an overview of the research background and context. It begins by establishing the study's contextual background, focusing on the status of English in Saudi higher education and PYP, and outlines the role of English and the structure of foundational preparation. One subsection discusses English in Higher Education and Preparatory Programmes. Another subsection explores English as a Medium of Instruction in engineering programmes in Saudi Arabia, focusing on policy and implementation within the discipline. The chapter also provides an overview of the concept of technical vocabulary and discusses the importance of vocabulary studies in English education in Saudi Arabia. As this research focuses on the use of technical vocabulary, it further describes the structure of the thesis. Additionally, the chapter offers an overview of the study and highlights the researcher's interest, personal motivation, and research focus. Chapter 1 concludes with the research objectives and research questions.

Chapter 2 reviews and discusses various theoretical and conceptual issues relevant to this thesis. It covers aspects of vocabulary knowledge, lexical profiling, and vocabulary load. Additionally, it offers a review of word lists and the history of developing various word lists in the literature through a corpus-based approach. The chapter emphasizes the importance of word lists in L2 learning, highlighting different types of word lists and the selection criteria used to generate them. Furthermore, it presents an overview of the disciplinary approach to written discourse, alongside a discussion of corpus-based lexical analysis related to written academic genres. The chapter concludes with a review of teacher perceptions of vocabulary learning and a summary of the key points discussed.

Chapter 3 presents an overview of the research methodology. It begins by explaining the research approach, whereby it shows that the present thesis is divided into three studies, which are based on the aforementioned eight research questions. In addition, in each study, the data collection and analytic approach adopted to answer particular research questions is described.

Chapter 4 presents the results and discussion of Study 1. Using AntWordProfiler, the results of the profiling analysis and vocabulary load are presented to address the first and second research questions. Research question 1 has a two-fold aim: first, to examine the lexical profiling of the EMDC related to Nation's BNC/COCA frequency lists (0 to 25,000) plus the supplementary list. Second, it examines the distribution of Schmitt and Schmitt's (2014) high, mid-, and low-frequency vocabulary in the EMDC (see Section 4.3.2).

Research question 2 focuses on assessing the vocabulary load in the EMDC. This is achieved by calculating the lexical coverage of each BNC/COCA frequency-based word list until it reaches either the minimum threshold of 95% or the optimal threshold of 98% coverage (Laufer, 1987; Van Zeeland & Schmitt, 2012).

Building on the analysis of lexical profiling and vocabulary load in the EMDC, Chapter 5 introduces Study 2, which centres on identifying the ETV, including both single-word and multiword units, in the EMDC. The chapter begins with an explanation of the methodologies employed in the study, including corpus-based and semantic approaches.

Next, the chapter presents the findings related to the three research questions outlined in Study 2. First, it addresses research question 3, which focuses on identifying single-word ETV items used in the EMDC and examines their distributions in Schmitt and Schmitt's (2014) high-, mid- and low-frequency vocabulary bands, respectively. Second, it presents research question 4, which focuses on identifying multiword technical vocabulary items found in the EMDC. Finally, it addresses research question 5, which aims to analyse the distribution of the ETV list across the five main sections of engineering masters' dissertations: Introduction, Literature Review, Methodology, Results and Discussion, and Conclusion.

Chapter 6 discusses Saudi undergraduate students' receptive knowledge of ETV developed in Study 2. It also discusses teachers' perceptions regarding the pedagogic usefulness of ETV lists. The chapter begins with the methodology used in the study. It first describes both learner and teacher participants. It also outlines the procedures for data collection from both learner and teacher participants. The UVLT (Webb, Sasao & Ballance, 2017) is administered to evaluate students' vocabulary proficiency, while a Yes/No receptive test is utilized to measure their receptive knowledge of ETV. The second part of this chapter intends to explore engineering lecturers' perceptions of the pedagogic usefulness of the ETV list in the engineering discipline. Then, it presents the results and discussion to address the research questions posed in this study.

Finally, Chapter 7 summarizes the key findings regarding all of the research questions posed in the thesis. It also outlines the study's contributions, implications, and limitations, along with concluding remarks. The chapter ends by offering recommendations for future research directions.

Chapter 2 : Literature Review

2.1 Introduction

This chapter reviews key conceptual and empirical studies that inform the central questions of the thesis. It first addresses vocabulary knowledge, with a focus on receptive and technical vocabulary, both of which are essential for understanding the lexical demands of specialised academic texts. The concept of vocabulary load, or threshold, is also examined. This refers to the proportion of known words required for adequate reading comprehension, often estimated between 95% and 98% in corpus-based studies (Coxhead & Demecheleer, 2018; Durovic, 2021).

The chapter then explores the development and role of word lists in vocabulary research. It covers various types, including general service lists (e.g., West's GSL and the NGSLs by Brezina & Gablasova, 2015, and Browne et al., 2013), academic word lists e.g., Coxhead's Acdemic Word List (AWL) and Gardner & Davies' Academic Vocabulary List (AVL), and discipline-specific lists, including those focused on engineering and multi-word items. The discussion also includes the criteria and counting units used in the construction of these lists.

Further attention is given to disciplinary discourse and how vocabulary use varies across academic genres. Studies have shown that lexical choices differ across research articles (Chang & Kuo, 2011; Jiang & Hyland, 2015; Swales, 1990), master's theses (Charles, 2003), and undergraduate writing (Lancaster, 2016). Finally, the chapter examines teacher perceptions of vocabulary learning, which are particularly relevant to this study's focus on engineering masters' dissertations written by Saudi students.

2.2 Key Aspects of Vocabulary Knowledge

This section discusses the aspects of vocabulary knowledge, emphasizing what it means to fully understand a word, as the present study seeks to explore students' knowledge of technical vocabulary. Vocabulary serves as a fundamental component of language, and assessing learners' proficiency in this area is essential (Schmitt, et. al, 2001). Vocabulary knowledge has long been recognized as a critical aspect of language teaching and learning. These include aspects such as learners' attitudes toward learning it and the strategies employed to acquire it (Trang, Truong, & Ha, 2018). Vocabulary knowledge is a multifaceted and crucial aspect of vocabulary research (Dang, Webb & Coxhead, 2022a; Schmitt, 1998; Nation, 2016). It encompasses both 'vocabulary size', which refers to the proportion of words known by a learner and 'vocabulary depth knowledge', which refers to the extent to which these words are understood. However, these two components are interconnected and differ in how each aspect is conceptualized and measured (Schmitt, 2010).

While researchers may conceptualize vocabulary knowledge in different ways, there is broad consensus that understanding a word involves much more than just grasping the connection between its form and meaning (Schmitt, 2010). It also encompasses aspects such as its discourse function, usage limitations, phonetic and syntactic features, and multiple meanings (Richards, 1976; Schmitt, 2010). Building on this, Nation (2013) developed a comprehensive framework for vocabulary knowledge, encompassing three key aspects: word form, meaning, and usage. Each of these aspects incorporates both receptive and productive knowledge (Nation, 2013). Examining learners' vocabulary knowledge can provide insights into their exposure to the target language (Dang, 2017; Schmitt, 2010).

2.2.1 Receptive Vocabulary Knowledge

The present study intends to explore learners' receptive knowledge of technical engineering terms. This section discusses this key aspect of vocabulary. Receptive vocabulary knowledge involves recognizing a word's form during listening or reading and recalling its meaning. On the other hand, productive vocabulary knowledge entails generating the word's meaning through speaking or writing and accurately recalling its form (spoken or written) (Nation, 2013). Schmitt (2010) defines receptive vocabulary knowledge as the ability to recognize and understand the meaning of words when encountered in listening or reading, without necessarily being able to produce them in speaking or writing (Schmitt, 2010).

A popular receptive vocabulary test developed by Nation (1983, 1990), known as the Vocabulary Levels Test (VLT), has been used to assess the receptive vocabulary knowledge of learners in L2/EFL contexts, such as Vietnam (Dang, 2020b), Iran (Akbarian, 2010), and Japan (Sakata, 2019). The VLT aims to measure the breadth of vocabulary known by second language (L2) learners in general English and has been further updated by scholars such as Webb, et al., (2017). Various studies (e.g., Nation, 2013; Schmitt, 2010; Webb & Rodgers, 2009a, 2009b) have shown that it is generally simpler to acquire and use a word receptively than to do so productively. The present study aims to investigate Saudi EFL learners' receptive knowledge of engineering technical vocabulary in the EMDC.

2.2.2 Technical Vocabulary Knowledge

Building on the conceptualization of technical vocabulary discussed in Chapter 1 (Section 1.6), this section shifts focus to the knowledge of such vocabulary within specialised fields. Technical vocabulary constitutes a significant portion of the lexicon in specialised texts (Benson & Coxhead, 2022; Dang, Coxhead, & Webb, 2017; Lu, 2018). However, as clarified in Section 1.6, it is a common misconception that technical vocabulary consists solely of rare or obscure words. In reality, technical vocabulary can include words from high-, mid-, or low-frequency levels (Nation, 2016). These words can appear across the BNC/COCA frequency lists (Coxhead & Demecheleer, 2018), a point often overlooked in earlier EAP studies (e.g., Coxhead, 2000; Sundayana & Kwary, 2019).

In this thesis, technical vocabulary knowledge refers to understanding the specialised meaning of a word beyond its general usage. As emphasized in ESP, language proficiency requires not only general linguistic knowledge of words but also subject-specific expertise or understanding of the technical meaning of words (Nation, 2016). Similarly, comprehending a technical word involves more than just general vocabulary knowledge; it also requires familiarity with the disciplinary concepts tied to the word (Coxhead, 2018; Hinkel, 2018). Developing technical vocabulary knowledge in a specialised field is crucial for learners (Dang, et al., 2022a). However, there is a limited understanding of the size of this lexical domain and how it expands over time (Benson & Coxhead, 2022). Learners must develop ways of thinking and frameworks of reference that align with the discourse of their field (Hirvela, 1997). To fully grasp a technical word, learners need to integrate both general vocabulary knowledge and the subject-specific knowledge relevant to that word within its technical context.

In ESP classes, teachers might need to explicitly teach specialised vocabulary to ensure students grasp the knowledge of important terminologies in their subject area. Due to the accuracy required for communication, productive word acquisition is also more challenging than receptive word learning (Nation, 2013). For this reason, L2 students benefit from the time spent in language classes learning the productive meaning, form, and context of technical vocabulary (Benson, & Coxhead, 2022). This research focuses on the knowledge of ETV by a group of Saudi engineering students.

2.3 Lexical Frequency Profiling

Lexical Frequency Profiling (LFP) is an analytic approach used for lexical analysis in vocabulary research, introduced by Laufer and Nation in (1995). It is used to assess the lexical richness and density of texts (Stamatović, Bratić, & Lakić, 2020). Over the past three decades, LFP has been extensively employed in numerous word list studies and has become a standard measure of productive vocabulary knowledge (Higginbotham & Reid, 2019). Laufer and Nation (1995) initially used LFP to analyse 65 short essays written by L2 learners by cross-referencing them with Praniskas' (1972) University Word List and West's (1953) GSL. They suggested that for lower-proficiency learners, vocabulary should be classified into three groups: the first 1,000 most frequent words, the second 1,000 words, and all other vocabulary. For advanced learners, the analysis expanded to include academic vocabulary, the second 1,000 words, and less frequent words that fell outside these categories (Gregori-Signesa & Clavel-Arroitiab, 2015). This approach aimed to provide a structured framework for assessing vocabulary knowledge based on frequency and proficiency levels.

In general, LFP is based on a corpus linguistics approach, and various computer programs have been developed for that purpose *e.g.,* Nation's *Range,* Anthony's *AntWordProfiler,* Webb's *Vocabprofile.* These are the most commonly used free software vocabulary profiling corpus-driven tools (Anthony, 2020). These programs, known as lexical profilers, can be used to generate vocabulary lists and calculate the amount of coverage in the corpus loaded for each of these lists. However, LFP has not been without criticism. For instance, Meara and Bell (2001) argue that the LFP approach does not accommodate multi-word units. There are different views on the unit of counting criteria used in generating word lists. These include counting tokens, lemmas, flemmas, word types, and word families within the target corpus (Dang, 2020; Lu, 2018; Therova, 2020) (see Section 2.4.1). These differences highlight the diverse approaches to defining and measuring lexical units in vocabulary analysis.

2.3.1 Vocabulary Load

This section discusses the relevance of vocabulary load in corpus-based vocabulary research. The current study aims to apply LFP to determine the lexical load of academic masters' dissertations in engineering, an academic genre that remains underexplored. In recent years, a growing area of research has focused on analysing different types of academic texts to address a key question about lexical load or threshold, which refers to the number of words learners need to know to effectively comprehend such texts (Coxhead, 2021; Coxhead & Demecheleer, 2018; Dang, 2020). The vocabulary load of a text plays an essential role in determining whether it is appropriate for learners at a particular level (Webb & Nation, 2008). Vocabulary load is determined by the concept of 'coverage', which measures the percentage of words in the text that learners need to know to achieve sufficient comprehension (Nation, 2013). Typically, corpus-based vocabulary load studies aim to identify the number of words required to reach sufficient reading comprehension, with coverage thresholds generally set between 95% and 98% (Durovic, 2021; Van Zeeland & Schmitt (2012).

Different perspectives have been put forward by vocabulary scholars (e.g., Coxhead, 2000; Nation, 2006; 2013; 2016; Nation & Webb, 2011) to clarify the vocabulary threshold that learners should acquire to sufficiently comprehend language used in speaking and writing (Laufer & Ravenhorst-Kalovski, 2010). Nation (2013) recommends using 'word families' as a unit of counting, which consist of a headword along with its derivatives and inflections, such as *write, writes, writing, writer, rewrite, unwritten, writable, written*, and *writability*, for counting words in receptive tasks and vocabulary load analysis. This approach takes into account the learning burden, defined as "the amount of morphological knowledge learners are expected to possess" (Nation, 2016, p. 9). However, some studies adopt 'lemmas' as the unit of counting e.g., Gardner & Davies, (2014), which include only the inflections of the headword, or count each word individually. Nation's (2012) BNC/COCA word frequency lists have been extensively utilized in corpus-based vocabulary load analyses, as also employed in the present study to find the vocabulary load of the EMDC.

Previous studies adopting Nation's (2012) BNC/COCA frequency list, such as Hsu (2018), found that knowledge of the first 10,000 BNC/COCA word families, plus proper nouns, ensures 98% lexical coverage in a traditional Chinese medicine (TCM) corpus. Similarly, Coxhead and Boutorwick (2018) demonstrated that nearly 98% coverage of Grade 8 Math textbooks requires familiarity with 8,000-word families, along with proper nouns, compounds, and abbreviations.

Sun and Dang (2020) found that 3000-word families are required to reach coverage of 95% of the vocabulary in English language textbooks used in Chinese high schools, while 9000-word families are required to reach 98% coverage. Ward (1999) found that a corpus of foundation engineering requires 2,000-word families to achieve a minimum of 95% coverage. Nation (2006) reported that 8000-word families, plus proper nouns, were required to cover 98% of the vocabulary load in a corpus of university textbooks, novels, and newspapers. Similarly, Coxhead et al. (2016) found that 8000-word families, plus supplemental lists of abbreviations, are needed to cover 98% of the vocabulary load in a specialised corpus of Carpentry.

Hsu (2014) found that around 5,000-word families are needed to achieve 98% coverage in engineering textbooks. In another study, Hsu (2013) found that around 14,000-word families, along with proper nouns and specialised medical terms, are required to reach 98% optimal coverage in medical textbooks. Dang and Webb (2014) found that 4,000-word families provide 95% coverage, while 8,000 are needed to achieve 98% coverage in spoken English lectures and seminars. Coxhead et al. (2017) revealed that 3,000-word families suffice for 95% minimal coverage in laboratory sessions, with 7,000 required for 98% optimal coverage. These findings suggest that written academic texts typically demand a more extensive vocabulary than spoken texts, though additional multidisciplinary research is necessary to confirm this trend (Coxhead, 2021). In line with this, the current study seeks to analyse the vocabulary load of engineering masters' dissertations written by Saudi students, using Nation's (2012) BNC/COCA word lists as the framework.

2.4 Word Lists

This section presents an overview of word lists, which are essential in vocabulary research, particularly in academic and technical fields like engineering. A word list is a collection of single-word and multi-word items organized based on frequency, range, or specialised function, designed to support students' learning objectives in general subjects or discipline-specific contexts (Benson, 2020; Dang, 2020). While earlier developed lists, such as the GSL (West, 1953) and the AWL (Coxhead, 2000), have guided vocabulary instruction, specialised lists like the Medical Word List (Hsu, 2013) and the Engineering List (Ward, 1999) address the unique needs of specific disciplines.

The use of word lists has immensely contributed to vocabulary learning in second language acquisition (Hunt & Beglar, 2005; Durrant, 2016; Therova, 2020). These lists are grounded in the assumption that certain words are more valuable for learners and that their usefulness can be predicted in advance (Durrant, 2016). Researchers like Coxhead (2000) further suggest that students with academic aspirations should prioritize general vocabulary before moving on to academic words, as the combination of these two categories forms the essential foundation for effectively comprehending academic texts. The following section examines the key units of counting employed in developing word lists for various purposes, providing insights into the most appropriate methodology for this study.

2.4.1 Overview of Key Counting Units Used in Word List Development

There is an ongoing debate in lexical profiling research regarding the most appropriate unit of counting for word list development (Coxhead, 2000; Gardner & Davies, 2014; Ha & Hyland, 2017; Wood & Webb, 2020; Vilkaitė-Lozdienė, Schmitt & Webb, 2020; Stamatović, et al., 2020). However, in the creation of single-word lists, five primary units of counting are commonly used: tokens, word types, lemmas, flemmas, and word families (Dang, 2020; Lu, 2018; Therova, 2020). Each of these units offers distinct advantages and challenges, shaping the ongoing discussion about the most effective approach to word list development.

A 'token' refers to the total number of individual word occurrences in a text or corpus, including repeated instances of the same word form (Nation, 2013). For example, if the word '*analyse*' appears three times in a text, it is counted as three separate tokens. Tokens are considered the most basic unit of word counting and provide a straightforward measure of word frequency (Lu, 2018).

On the other hand, a 'lemma' groups together a base word e.g., *analyse* and its inflected forms e.g., *analyses*, *analysed*, *analysing*, all of which belong to the same word class and are treated as a single unit (Dang, et al., 2017; Francis & Kučera, 1982). This approach simplifies word counting by consolidating related forms, making it particularly useful for understanding word usage patterns in texts.

A 'word family' e.g., *compute* consists of a base word or stem *compute*, its inflected forms *computes, computed, computing*, and closely related derivations up to Bauer and Nation's (1993) Level-6. This includes forms such as *computable, incomputable, computability, computation, computational, computationally, computer, computers,* and *computerized*. Word families have traditionally been used as the counting unit in many older general academic word lists, such as Coxhead's (2000) AWL and Xue and Nation's (1984) work. They remain the standard unit of counting in numerous general vocabulary lists (Dang, 2020; Lu, 2018; Liu & Han, Therova, 2020; 2015; Yang, 2015). This approach groups morphologically related words together, providing a broader yet cohesive representation of vocabulary knowledge.

Researchers such as Brezina and Gablasova (2015), Gardner and Davies (2014), and Lei and Liu (2016) have raised concerns about the use of word families as a unit of counting in word list development, suggesting that lemmas are a more appropriate alternative. Nation (2016) offers a comprehensive rebuttal to these criticisms. One major concern is that word families may include members with meanings that are not closely related, making them less precise for counting compared to lemmas (Gardner & Davies, 2014). In response, Nation (2016) emphasizes that Bauer and Nation's (1993) scale ensures semantic consistency by including only words whose meanings remain closely tied to the base word, whether standing alone or in derived forms.

Bauer and Nation (1993, pp. 257-262) introduced a seven-level framework for categorizing word families based on their affixation patterns and complexity. At Level 1, words are stand-alone base forms without affixes, such as *create*. Level 2 incorporates regular inflections, including plurals, verb tenses, comparatives, superlatives, and possessives, resulting in forms like *creating*, *created*, and *creates*. Level 3 focuses on common and regular derivational affixes, such as *-er, -less, -ly, -ness, -ish* and *able*, producing words such as *childish* and *removable*. Level 4 expands to include frequent and orthographically regular affixes, such as *-al, -ess, -ism, -ist, -ity, -ment,* and *-ous,* for example, the words *ritualism* and *ritualist*. Level 5 covers less frequent but regular affixes, including a wide range of prefixes and suffixes, such as *-age,* and *-hood,* for example, and words like *stoppage* and *brotherhood*. Level 6 involves frequent but irregular affixes, such as *-ic, -pre, re-,* and *-ify,* forming words like *dramatic* and *premature*. Finally, Level 7 includes classical roots and affixes, such as *ab-, ad-, com-, de-,* *dis-, ex-* and *sub-,* often derived from Latin or Greek, used to construct more complex word families, for example, *declassify, submarine, disagree*, and *exclude*.

Another issue raised is that word families do not differentiate parts of speech, whereas lemmas do (Gardner & Davies, 2014; Lei & Liu, 2016). Nation (2016) acknowledges this limitation but argues that part-of-speech distinctions, while useful for differentiating closely related items like *walk* (verb) and *walk* (noun), fail to address homonyms with the same part of speech, such as *bank* (financial) and *bank* (river).

An alternative approach to word counting in corpus analysis involves the use of the *lemma*, defined as a unit that "distinguishes part of speech and includes the stem of the word and its inflectional suffixes" (Nation, 2016). The lemma has been widely used as the basis for developing word lists, such as the New General Service List compiled by Brezina and Gablasova (2015). It is often regarded as particularly appropriate for pedagogical word lists because it categorizes words according to their grammatical roles (Gardner & Davies, 2014). For example, the verb *cross* and its inflected forms (*crosses, crossing, crossed*) belong to one lemma, while the noun *cross* is classified as a separate lemma.

A further development of this approach is the *flemma* (or "frequency lemma"), which has been adopted in several recent studies targeting beginner learners (Browne, 2014; Dang & Webb, 2016; McLean, 2017; Pinchbeck, 2014). The flemma expands the lemma unit and is considered especially suitable for novice learners because it minimizes the need for morphological knowledge when acquiring vocabulary.

Despite their usefulness in general vocabulary research, neither the lemma nor the flemma is recommended for identifying technical vocabulary. This is because not all members grouped under a lemma or flemma carry technical meanings, which can lead to imprecise classification in discipline-specific word list construction (Chung & Nation, 2004; Nation, 2016).

Word types are therefore increasingly recommended for identifying technical vocabulary in specialised corpora (Coxhead & Lu, 2020; Liu & Lei, 2020). As in the present study, the unit of analysis for identifying technical words was *word type*, as research consistently demonstrates that it is the most suitable measure for counting highly specialised vocabulary (Chung & Nation, 2004; Coxhead et al., 2016; Ha, 2015; Nation, 2013; Nation et al., 2016; Ward, 2009). This is because one form of a word may hold a precise technical meaning, while its other forms do not. For example, Coxhead et al. (2016) found that *fixing* is a technical term in carpentry, whereas *fix* is not. Similarly, Ha (2015) showed that *hierarchy* is central in finance, while *hierarchical* and *hierarchies* are not. Moreover, knowledge of one form of a word does not necessarily imply knowledge of its derivatives or inflections (Durrant, 2013; Martinez & Murphy, 2011; Ward, 2009). For instance, Martinez and Murphy (2011) argued that learners are unlikely to deduce the meaning of *fishy* (suspicious) from *fish*, given the semantic distance between the two.

Accordingly, this study adopts *word type* as the unit of counting in identifying engineering technical vocabulary in the Engineering Masters' Dissertations Corpus (EMDC), ensuring that the analysis captures the precise lexical items essential for comprehension and production within the engineering discipline. For example, in engineering, the word type *torque* refers specifically to a fundamental mechanical concept, whereas its plural *torques* may appear infrequently or in less technical contexts. Likewise, *current* as a noun denotes an essential concept in electrical engineering, while *current* as an adjective (e.g., *current issues*) is not technical. By focusing on word types, researchers ensure that only the relevant technical terms are identified and prioritized, without conflating them with less specialised uses (e.g., Coxhead et al., 2016; Watson Todd, 2017). This approach aligns with the need to capture the precise vocabulary required for specialised fields. Therefore, the present study adopts the word type as the unit of counting in the identification of engineering technical vocabulary in the EMDC.

2.4.2 Overview of Selection Criteria Adopted in Word List Generation

This section discusses the selection criteria, which are critical factors to address when developing or assessing corpus-based word list research. The outcomes of corpus-based word list research can vary significantly depending on the methods used for counting words (Nation, 2016). Frequency, range, and dispersion are the most commonly used criteria in the objective corpus-driven approach for word list research (Brezina & Gablasova, 2015; Byrd & Coxhead, 2010; Dang et al., 2017; Dang, 2020). However, range and dispersion are now discouraged in discipline-specific word list research (Dang, 2020; Durovic, 2021; Lei & Liu, 2020; Lu & Coxhead, 2020).

The term 'frequency' refers to the total number of times a word or multiword unit appears in a corpus. Ranking words based on their frequency of occurrence is one of the most fundamental uses of corpora (Szudarski, 2018). Schmitt and Schmitt (2014) suggest categorizing vocabulary into three groups based on frequency analysis: 'high-frequency', 'mid-frequency', and 'low-frequency' words. While Schmitt and Schmitt (2014) propose that the first 3,000 words of Nation's (2012) BNC/COCA frequency lists should be classified as high-frequency vocabulary, it is widely accepted that high-frequency English vocabulary comprises the 2,000 GSL most common word families (Szudarski, 2018). The present study utilizes Schmitt and Schmitt's (2014) frequency classification, as adopted in various specialised vocabulary research studies such as Benson and Coxhead (2022) and Coxhead and Lu (2020).

Range refers to the number of different texts or sub-corpora in which a word appears. This criterion has been consistently used in research developing specialised word lists in ESP (Dang, 2020; Durovic, 2021; Lei & Liu, 2020; Lu & Coxhead, 2020). Its purpose is to ensure that students across various disciplines and subject areas can benefit from learning these word lists (Dang et al., 2017). A selected word family was required to appear in at least 50% of the topic areas and across all disciplinary sub-corpora (Dang et al., 2017).

Dispersion, on the other hand, refers to how evenly a word is distributed across different texts or sub-corpora. The greater a word's frequency, range, and even distribution, the more likely it is that learners will encounter it in their language use (Dang et al., 2017). While several methods exist for measuring dispersion, Juilland and Chang-Rodrigues'(1964) approach is the most widely used in research focused on creating both specialised and general word lists (e.g., Dang, 2020; Durovic, 2021; Gardner & Davies, 2014; Lei & Liu, 2016). The dispersion value in Juilland and Chang-Rodrigues' (1964) method ranges from 0, indicating "very uneven distribution," to 1, representing "perfectly even distribution."

In the present study, however, neither range nor dispersion was used as a selection criterion for identifying technical vocabulary in engineering masters' dissertations. This is because the technical status of a word in engineering is defined by its conceptual significance within the discipline, rather than by how evenly it is spread across texts. In a specialised genre such as dissertations, essential terms often appear in concentrated contexts, within specific chapters or projects, resulting in low range and dispersion values. Excluding such words on statistical grounds would risk overlooking core engineering vocabulary central to advanced research. However, range and dispersion are now discouraged more broadly in discipline-specific word list research. Consequently, several recent studies focusing on discipline-specific word lists did not include range and dispersion as selection criteria when identifying technical words (Benson & Coxhead, 2022; Dang, 2020; Durovic, 2021; Lei & Liu, 2020; Lu & Coxhead, 2020). These measures were also excluded from the current study for the following reasons.

First, the present study focuses on identifying technical vocabulary within the specific genre of master's-level engineering dissertations. For this purpose, range and dispersion were deliberately excluded as selection criteria. This decision is grounded in the understanding that a word's technical status is defined by its conceptual importance within the discipline, not its statistical distribution across a corpus. This principle is clearly illustrated by prior research in other highly specialised fields. For instance, in a study on Traditional Chinese Medicine (TCM) vocabulary, Lu (2018) demonstrates that key technical terms are inherently concentrated within their specific domain. This finding aligns with the argument of Miller and Biber (2015) that technical vocabulary is often localized, appearing frequently within a specific text or chapter that delves into a specialised topic. The nature of a master's dissertation—a deep, focused investigation into a narrow research problem—means that essential technical terms are highly likely to appear with low range and dispersion across the wider corpus. A compelling example from Fraser (2007) further confirms this rationale: the term *parasympathetic* was identified as a critical technical word in pharmacology despite being found in only three articles. Similarly, in an engineering dissertation corpus, core techniques, proprietary software names, or specific material properties vital to a subset of projects may appear in only a few texts. Applying range and dispersion criteria would systematically eliminate such precise, yet essential, vocabulary. Therefore, to ensure the resulting word list accurately captures the specialised lexicon of engineering research at the master's level, these statistical measures were not employed.

Second, Durovic (2021) argues that range and dispersion are more relevant to research involving large corpora used to extract generic academic words for multiple disciplines (e.g., Coxhead, 2000; Coxhead & Hirsch, 2007; Gardner & Davies, 2014) than for studies targeting one discipline-specific field (e.g., Coxhead & Demecheleer, 2018). Dispersion is a suitable criterion when identifying general academic word lists because it can show how evenly a word is distributed across a broad corpus. However, due to the concentrated nature of specialised corpora, it is not applicable when identifying discipline-specific word lists (Dang, Coxhead, & Webb, 2017). For this reason, the present study did not employ dispersion, as it focuses on identifying discipline-specific vocabulary in engineering masters' dissertations.

Finally, when creating discipline-specific word lists, criteria like range and dispersion can be misleading. As Dang, Coxhead, and Webb (2017, p. 5) argue, a word may be highly frequent not because it is useful across many fields, but because it is overused in a single subject. Consequently, profiling tools like Range or AntWordProfiler might incorrectly eliminate essential discipline-specific technical words due to an uneven distribution. For these reasons, range and dispersion were excluded as criteria in the present study, ensuring that the resulting list captures the most relevant and specialised vocabulary used in Saudi engineering masters' dissertations.

For multiword units, common selection criteria include keyword analysis, N-grams, mutual information, t-scores, and log-likelihood. These criteria assess the extent to which the components of a sequence co-occur more frequently than would be expected by chance (Benson & Coxhead, 2022; Coxhead et al., 2017; Dang, 2020; Durovic, 2021). Another criterion used to identify technical vocabulary, both single-word and multiword units, is semantic analysis. Chung and Nation's (2003) semantic rating scale has been adapted for numerous ESP studies (e.g., Hsu, 2014; Quero, 2015; Coxhead & Demecheleer, 2018; Coxhead et al., 2016). (For more details on selection criteria, see Chapters 3 and 5.) In the present thesis, corpus-based and semantic analytical methods, including frequency principles and specialised occurrences, were employed to identify engineering technical vocabulary items from the EMDC. These items were subsequently used to create lists of engineering single-word and multiword units for pedagogical use in university-level courses.

2.4.3 Types of Word Lists

Corpus linguistics and applied linguistics research have focused on developing vocabulary lists for educational purposes, particularly for EAP and ESP courses (Đurović, Stamatović & Vukicevic, 2021). This has led to the creation of several influential word lists, including West's (1953) GSL, Xue and Nation's (1984) University Word List, Coxhead's (2000) AWL, Browne et al.'s New Academic Word List (NAWL), and Gardner and Davies's (2014) AVL. There are three main types of word lists: general, academic, and specialised or discipline-specific word lists, as discussed in the following sub-sections.

2.4.3.1 General Service Lists

This section introduces the three general service lists, starting with West's (1953) GSL. As the first core high-frequency word list developed, the GSL has been widely recognized as a valuable resource for both teaching and testing purposes (Dang, 2020; Granger & Larsson, 2021; Tongpoon-Patanasorn, 2018). West's (1953) GSL consists of 2,000 high-frequency word families developed from a corpus of 5 million written texts (Lu & Coxhead, 2020). The list constituted core lexicons relevant to any language use, including general communication, academic communication, and professional communication, (Coxhead, 2018). Given the age of the corpus from which this list was created, it is currently considered fairly outdated (Coxhead, 2000; Dang et al., 2017). In addition, the GSL was manually compiled from a pre-computer era corpus, which is considered small by modern standards for creating comprehensive word lists. Another limitation is the lack of a clear definition of what constitutes a ‘word’ (Browne, 2013). Despite its limitations, West’s (1953) GSL continues to hold relevance in vocabulary research, as it covers around 80% of the words commonly found in most texts (Csomay & Prades, 2018).

To address the limitations posed by the GSL, two groups of researchers developed updated versions: Brezina and Gablasova (2015) created the new-GSL, and Browne, Culligan, and Phillips (2013) introduced their own revised version of NGSL. These updates aimed to modernize and refine the original GSL to better meet the needs of contemporary language learners. Both versions of the New-GSL were built from large corpora. Brezina and Gablasova (2015) developed their new-GSL from a 12-billion-word corpus, while Browne et al. (2013) created their NGSL from a 273-million-word corpus (a subset of the Cambridge English Corpus). Both versions report slightly greater coverage than the GSL. The NGSL consists of 2,494 lemmas and provides coverage of 80.1% to 81.7% of the content in the source corpora.

On the other hand, Browne et al. (2013) created their own version of the NGSL from a 273-million-word subset of the 2-billion-word Cambridge English Corpus (CEC). This version of the NGSL contains approximately 2,800 high-frequency words. It uses a modified lexeme approach, which counts the headword across all parts of speech and includes all inflected forms, differing from the traditional lexeme definition that groups related forms with the same meaning and part of speech (Browne, 2014).

2.4.3.2 Academic Word Lists

This section discusses the development of academic word lists. Academic vocabulary, or academic word lists, refers to words outside the GSL that are frequently found in a wide range of academic texts, both written and spoken, specifically identified to support general academic learning (Coxhead, 2016, 2020; Hyland, 2016; Watson Todd, 2017). Farrell (1990, p. 11) defined academic vocabulary as "formal, context-independent words with a high frequency and/or wide range of occurrence across scientific disciplines, not usually found in basic general English courses". However, there is ongoing debate about the existence of a universal core academic vocabulary (Hyland & Tse, 2007). Coxhead (2016) emphasizes that words common in one discipline may not share the same meaning or usage in another.

The early academic word lists were constructed by analysing student annotations of unfamiliar words in textbooks, with frequency serving as the primary selection criterion. For instance, Lynn (1973) created a 197-word family list from 10,000 notes taken from 52 books, while Ghadessy (1979) generated lists using a 478,700-token corpus from 20 textbooks across three disciplines in Iran. Xue and Nation (1984) later consolidated these earlier lists into the University Word List (UWL), which inherited limitations due to the absence of standardized principles. Despite these shortcomings, the UWL achieved significant lexical coverage in academic texts – 8.5% (Xue & Nation, 1984) and 9.8% (Coxhead, 2000) – and remained widely used until Coxhead's (2000) AWL was introduced. These early academic lists were constrained by the limited text collections available at the time, as technological barriers hindered large-scale corpus development (Dang, 2017). These limitations led to the development of modern and comprehensive academic word lists, which are generally based on corpora of academic texts (Coxhead, 2000; Coxhead, 2016; Coxhead & Nation, 2001; Dang, et. al, 2017).

Coxhead's (2000) AWL is the most widely recognized among modern academic word lists. It was created from a 3.5-million-word corpus of multiple academic disciplines. It comprises 570-word families after excluding West's (1953) GSL and covers roughly 10% of academic texts. Despite its widespread use, the AWL has notable limitations, such as its reliance on word families, which prompted the creation of more refined and comprehensive lists (Gardner & Davies, 2014; Therova, 2020).

To address the selection criteria limitation in AWL, Gardner and Davies (2014) created their AVL, using lemmas instead of word families to provide better insights into word meanings and functions. The AVL consists of 3,015 lemma headwords and 1,991-word families, covering approximately 14% of a 120-million-word corpus. Subsequent research efforts have produced more alternative lists with larger corpora and refined selection criteria. For example, Browne et al. (2013) developed the New Academic Word List (NAWL), consisting of 960 words based on a 288-million-word corpus, which includes three main parts: the CEC Academic Corpus, containing 248 million words (86.3%); academic textbooks, comprising 36 million words (12.6%); and the Oral Corpus, which includes 3 million words (1.1%) (Therova, 2020).

Acknowledging the differences between spoken and written academic English, Dang et al. (2017) developed the Academic Spoken Word List (ASWL) to address learners' comprehension of academic speech. The ASWL consists of 1,741-word families graded based on four levels of Nation's (2012) BNC/COCA lists. The ASWL covers 90.13% of a 13-million-word corpus of academic spoken discourse. The criteria used by the authors to create the ASWL include a frequency cut-off point of 350 and a dispersion cut-off point of 0.6, which represent the optimal balance among the four ASWL features. These thresholds are lower than those used for Coxhead's (2000) AWL (frequency of 370) and Gardner and Davies's (2014) AVL (dispersion of 0.8), reinforcing previous research findings that academic speech and academic writing exhibit distinct linguistic characteristics (Biber, 2006; Biber et al., 2002).

In recent years, vocabulary research has increasingly focused on discipline-specific word lists to meet the unique language demands of various fields. Building on this trend, the present study aims to identify technical vocabulary in engineering masters' dissertations written by Saudi students, with the goal of creating an engineering-specific word list for pedagogical purposes. The findings are expected to provide valuable insights for educators, helping them evaluate the relevance and usefulness of technical vocabulary in engineering discourse.

2.4.3.3 Discipline-Specific/Specialised Word Lists

This section introduces ‘discipline-specific word lists’, also referred to as specialised word lists (Benson, 2020; Lu, 2018). Dang (2020) describes a ‘discipline-specific word list’ as a collection of words that occur within a particular discipline or across a range of related academic subjects, designed to support learning in ESP. These lists can differ in scope, with some being broad lists that encompass entire fields (e.g., Kwary, Ratri & Artha, 2017) and others being narrow lists that target specific subfields (e.g., Coxhead, 2018; Yang, 2015). Some narrow discipline-specific word lists are categorized as ‘low-frequency vocabulary lists’, consisting of words that appear infrequently across a wide range of texts. However, Nation and Kyongho (1995) point out that there is no consensus on whether low-frequency words are inherently tied to a narrow range or represent technical vocabulary from other disciplines, as what is considered technical vocabulary for one person may be a low-frequency vocabulary for another. Consequently, the size and coverage of these word lists depend on the purpose and criteria used in their creation. Some lists are built ‘from scratch’, without excluding general words, while others are developed by first removing words from the most common general vocabulary and, in some cases, academic vocabulary (Đurović, et al., 2021).

In a more specific context, another type of specialised word list is referred to as subject-specific word lists, also known as stand-alone discipline-specific lists, which concentrate on a single disciplinary discourse (Dang, 2020; Lu, 2018; Lu & Coxhead, 2020; Liu & Lei, 2019). These lists are particularly relevant for ESP programmes where all learners are preparing to study the same subject area, often referred to as ‘technical vocabulary’ (Liu & Lei, 2019). Hence, discipline-specific word lists are better suited to meet the needs of such learners, whereas general academic word lists offer limited assistance to those with highly specialised requirements in a single field (Dang et al., 2017; Durrant, 2016). These lists emphasize the shared vocabulary within a specific subject area, such as the Carpentry List (Coxhead et al., 2016). Proficiency in technical vocabulary is also crucial for success in ESP programmes (Knoch, 2014). Given the subject-specific nature of these lists, they vary significantly from one field to another. Moreover, technical vocabulary is pervasive and highly frequent in professional language, making it an invaluable resource not only for students studying the subject but also for professionals working directly or indirectly in the field (Liu & Lei, 2020).

Two main approaches are used to develop discipline-specific word lists for targeted subject areas. The first approach depends on existing generic high-frequency word lists, such as the University Word List (UWL) by Xue and Nation (1984) and the AWL by Coxhead (2000). The second approach, however, creates stand-alone lists without relying on any existing word lists. While the former method considers learners' prior knowledge of general high-frequency vocabulary, it is constrained by the limitations of the broad word lists it builds upon (Dang et al., 2017). Additionally, this approach risks including lexical items that learners may already be familiar with (Dang et al., 2017). Table 2.1 outlines some discipline-specific word lists derived from these two methodologies.

Table 2-1:Some of the Discipline-Specific Word Lists (Single Word Only)

Author(s)	Name of the word list	Corpus size (tokens)	Source	Size of the list	Selection criteria
Coxhead and Hirsh (2007)	Science Word List (SWL)	1,761,380	Sciences course books and materials suggested by lecturers	318-word families	Frequency, range, and dispersion Exclusion of GSL and AWL
Wang et al. (2008)	Medical Academic Word List (MAWL)	1,093,011	288 Medical RAs	623-word families	Frequency, range, and specialised occurrence Excluding GSL but including AWL

Hsu (2013)	Medical Word List (MWL)	15,000,000	155 medical textbooks (31 medical subject areas)	595-word families	Specialised occurrence, range and frequency of a word family Excluding BNC 3000 but including AWL
Liu and Han (2015)	Environmental Academic Word List	862,242 words	A corpus of 200 RAs (from the Science Citation Index)	458-word families	Frequency, range, exclusion of GSL
Yang (2015)	Nursing Academic List	1,006,934 words	252 Nursing RAs from online sources (11 subject areas)	676-word families	Range, frequency, and word family comparison Excluding GSL but including AWL
Lei and Liu (2016)	New Medical Academic Word List	Two corpora: 2,700,000-words (MAEC) 3,500,000-words (MTEC)	RA (Medical journal articles) and textbooks	965 lemmas	Frequency, ratio, dispersion, specialised occurrence, Chung and Nation's (2003) semantic rating scale, excluding GSL

MAEC: Medical Academic English Corpus, MTEC: Medical English Textbooks Corpus

It is important to acknowledge that scholars employ varying principles when selecting specialised word lists, depending on the intended purpose of the list, as discussed in previous sections. The studies reviewed in Table 2.1 illustrate this diversity. For instance, the early discipline-specific word list developed by Coxhead and Hirsh (2007) adopted similar selection principles – 'frequency', 'range', and 'dispersion' – as used by Coxhead (2000) in creating her AWL. However, they later excluded the GSL and AWL from their criteria. Similarly, Liu and Han (2015) developed the Environmental Academic Word List using frequency and range while excluding the GSL, but they did not consider dispersion. Meanwhile, Hsu (2013) developed the Medical Word List (MWL), and Yang (2015) created the Nursing Academic Word List (NAWL), both following three principles: range, frequency, and word family comparison. These lists excluded the GSL but included the AWL. Similarly, Lei and Liu (2016) created their New Medical Academic Word List by analysing two corpora: a 2,700,000-word Medical Academic English Corpus (MAEC) and a 3,500,000-word Medical English Textbooks Corpus (MTEC), which included research articles from medical journals and textbooks. They developed a list of 965 lemmas based on criteria such as frequency, ratio, dispersion, semantic analysis, and specialised occurrences while excluding words from the GSL.

However, the size of the lists depends on the discipline and genre, and sometimes the size of the corpora, in addition to the principles of selection. As illustrated in Table 2.1, the corpora were compiled from textbooks, course materials, and research articles. It shows that theses or dissertation genres are underexplored. Therefore, the aim of the present study is to contribute to research on dissertation genres by examining technical vocabulary knowledge and its usage in Saudi engineering Masters' dissertations. The following table outlines studies focused on subject-specific word lists.

Table 2-2:Some of the Subject-Specific Word Lists (Single Word Only)

Author(s)	Discipline-specific word list	Corpus size	Source	Size of the list	Selection criteria
Fraser (2009)	Pharmacology Word List	360,000-running words	RAs	2000-word families	Including both GSL and AWL
Valipouri and Nassaji (2013)	Chemistry Academic Word List	4,000,000 running words	1,185 Chemistry (RAs)	1,577-word families	Frequency, range, and specialised occurrence
Coxhead, Demecheleer, and McLaughlin (2016)	Carpentry List	300,594 running words	Carpentry instructional manuals and workbooks	1,424-word types	Frequency principles (adopting Nation's 2012, BNC/COCA), Chung and Nation's (2003) semantic rating scale,
Coxhead and Demecheleer (2018)	Plumbing Word List	565,881 words (written corpus) and 133,093 words (a spoken corpus)	Plumbing instructional manuals and workbooks (written corpus)	1,456-word types	Frequency principles (adopting Nation's 2012, BNC/COCA) Chung and Nation's (2003)

			18 hours of recordings of lectures and practical sessions (spoken corpus).		semantic rating scale
Hsu (2018)	TCM English Word List	13,000,000 running words	TCM textbooks	605-word families	Frequency principles (adopting Nation's 2012, BNC/COCA) Chung and Nation's (2003) semantic rating scale, Range, dispersion, and keyword analysis
Lu (2018)	TCM technical word lists	1,171,625 running words (theoretical) 1,109,701 running words (practical) 1,000,000 running words (RAs)	TCM textbooks (theoretical and practical) and research articles (RAs)	2,778-word types	Frequency principles (adopting Nation's 2012, BNC/COCA). Chung and Nation's (2003) semantic rating scale; keyness analysis
Tongpoon-Patanasorn (2018)	Technical Words List for Finance	2,004,964 running words	Textbooks, RAs, newspapers, and websites	569-word families	Frequency principles, Chung and Nation's (2003) semantic rating scale, keyness analysis

TCM = Traditional Chinese Medicine

Table 2.2 provides a summary of subject-specific word lists (single-word items) developed across various disciplines, including sciences: chemistry (Valipouri & Nassaji, 2013), pharmacology (Fraser, 2009), Chinese traditional medicine (Hsu, 2018), and Lu (2018) in technical sciences: carpentry (Coxhead, et al., 2016), plumbing (Coxhead & Demecheleer, 2018). Other subjects include finance (Tongpoon-Patanasorn, 2018).

These word lists were developed using single-word lexical items, mostly from written language, including textbooks, course materials, and research articles (e.g., Fraser, 2009; Valipouri & Nassaji, 2013; Yang, 2015). Two studies conducted by Coxhead and Demecheleer (2018), and Coxhead, McLaughlin, and Reid (2018) involved corpora compiled from both written and spoken language. Word families are used as the unit of counting lexical items in six of these word lists (Fraser, 2009; Tongpoon-Patanasorn, 2018, Valipouri & Nassaji, 2013; Yang, 2015), while word types are employed in studies by Coxhead et al. (2016), Coxhead, McLaughlin, and Reid (2018), Lu (2018), and Coxhead and Demecheleer (2018).

Frequency principles (mostly adopting Nation's 2012 BNC/COCA base lists) and semantic rating (adapted and modified from Chung & Nation, 2003) are the main selection criteria employed in these studies. Two studies (Valipouri & Nassaji, 2013; Yang, 2015) used range and frequency. While range and dispersion are common selection criteria for word lists, they are not used in these studies for several reasons, as highlighted by the researchers. For example, the specialised corpora are relatively small. Some of the material contained texts of different sizes, and some specialised words are opaque and may appear in small numbers in one text than in another (Coxhead, McLaughlin, & Reid, 2018). Notably, Lu (2018) developed the Traditional Chinese Medicine (TCM) Technical Word List, comprising 2,778-word types, from a corpus totalling 3,281,326 running words. This corpus included 1,171,625 words from theoretical-based TCM textbooks, 1,109,701 words from practical-based TCM textbooks, and 1,000,000 words from research articles (RAs). The selection criteria incorporated frequency principles based on Nation's (2012) BNC/COCA lists, Chung and Nation's (2003) semantic rating scale, keyness analysis, and a comparison with general written English corpora. The present study employs BNC/COCA base lists and incorporates keyness analysis, semantic rating scales, and word types as the unit of selection.

2.4.3.4 Discipline-Specific Lists of Multiword Units

While the majority of vocabulary research has historically focused on knowledge of individual words, there is now a growing emphasis on studying the acquisition of vocabulary beyond the single-word level (Coxhead, 2021). Multiword units are now widely accepted as part of the vocabulary learning curriculum, alongside single-word items (Pellicer-Sánchez, 2020). Beyond single words, another trend in vocabulary research is the development of lists of multiword items. Corpus research has identified a tendency for words to co-occur frequently and form multiword units, which are a crucial feature of naturally occurring language (Biber & Barbieri, 2007). This implies that, beyond analysing the frequency of individual words in various academic texts, corpus linguists have also examined the most common word clusters and investigated their roles in shaping the structure of academic discourse (Szudarski, 2018). Multiword units are essential in language because they can provide structural frames for language expressions (Biber et al., 2004), which are crucial for learning a language for academic purposes (Coxhead, Rahmat & Yang 2020; Durrant, 2016; Simpson-Vlach & Ellis, 2010).

However, word list research on multiword units is more limited in number compared to single-word lists (Coxhead, 2021). These lists are primarily focused on general academic multiword lists, such as the lexical bundles list by Biber et al. (2004), the Academic Collocation List (ACL) by Ackermann and Chen (2013), the Academic English Collocation List by Lei and Liu (2018), and the Academic Formulas List by Simpson-Vlach and Ellis (2010). However, there are few specialised multiword units in the literature, as presented in Table 2.3.

Table 2-3: Technical/Discipline-Specific Multiword Items

Author(s)	Discipline-specific word list	Corpus size	Source	Size of the list	Selection criteria
Wood and Appel (2014)	Multiword constructions (MWC)	1,600,000 million running words	10 most widely used EAP textbooks in the fields of business and engineering in Canadian universities	94 multiword units	Frequency analysis, range, keyword analysis using Scott's (2007) Wordsmith concordance tool
Ha (2015)	Finance Lexical Bundles	6,753,212 words	Annual reports and earnings call transcripts	539 technical lexical bundles	Frequency, range, and meaning criteria
Coxhead, Dang and Mukai (2017)	Multiword unit list (in university tutorials and laboratories)	137,399 (EAP/ESP textbook) 380,078 (Lab tutorials)	EAP/ESP textbook Lab tutorials and laboratories	176 phrases	Frequency and range principles (adopting Nation's 2012, BNC/COCA), keyword analysis
Benson and Coxhead (2022)	Multiword unit word lists (spoken rugby discourse)	61,295 words	Authentic Interaction Corpus (25,637) TV commentary corpus (35,658)	252 technical single words (12%) 267 multiword unit list	Frequency and range principles (adopting Nation's 2012, BNC/COCA), semantic rating scale (specialised occurrence)

Table 2.3 summarizes the unit of counting, corpus construction, and selection criteria of some specialised multiword units in different disciplines. Wood and Appel (2014) investigated multiword constructions (MWCs) in Business English and engineering textbooks, emphasizing their importance for EAP educators and materials developers. They created the First Year Business and Engineering Textbook Corpus (FYBETC), consisting of 1.6 million running words drawn from ten popular EAP textbooks used in business and engineering programmes at a Canadian university. Applying frequency and range as selection criteria, they identified

MWCs that occurred at least 40 times (equivalent to 25 instances per million words) and appeared in a minimum of two textbooks from each discipline. For comparative analysis, they also examined a 151,839-word corpus from contemporary EAP textbooks. Their findings revealed that multiword constructions (MWCs) were infrequently present in EAP textbooks for business and engineering at the Canadian university and were not addressed for pedagogical purposes.

Ha (2015) created a list of 539 technical lexical bundles by analysing a financial corpus containing 6,753,212 words sourced from annual reports and earnings call transcripts. These bundles, consisting of two to four words, were selected based on frequency, range, and meaning criteria. The author categorized these lexical bundles into two types: 'visible' or 'least technical' lexical bundles, where the meaning can be directly inferred from the individual words e.g., '*investment portfolio'*, '*cash flow analysis'*, and '*synergistic technical'* bundles, whose meaning is not immediately apparent from the individual words e.g., *'market capitalization*', '*interest rate swap'*. The 'visible' technical bundles were further distinguished from the 'least technical' bundles by the presence of at least one technical term. This classification highlights the varying degrees of technicality in financial language.

More recent studies focus on generating single and multiword items in one study. For example, Coxhead, Dang, and Mukai (2017) created technical single and multiword unit word lists from two corpora: one derived from tutorials and laboratories (380,078 words from recorded sessions of university laboratories and tutorials) and another from EAP/ESP textbooks (137,399 words). The authors developed 176 phrases using Nation's (2012) BNC/COCA lists, ranked according to frequency and range, with a frequency threshold of 25 occurrences per million words.

Benson and Coxhead (2022) developed both multiword units and single technical vocabulary in spoken rugby discourse. The authors compiled a corpus comprising 61,295 running words from spoken rugby texts. The analytic approach employed included vocabulary load analysis using Nation's (2012) BNC/COCA frequency lists, along with supplementary lists and semantic rating scales. The findings indicated that learners must be familiar with more than 4,000-word families, in addition to the supplementary lists, to attain 98% comprehension of the rugby corpus. Within this corpus, they identified 252 technical single words, which accounted for over 12% of the spoken rugby discourse. Additionally, they developed a pedagogically oriented list of 267 multiword units. To assess the knowledge of these terms, they administered a receptive knowledge test to 77 participants, including both English L1 and L2 speakers. The results revealed differences in technical rugby vocabulary between native and non-native English speakers. The study concludes with an examination of the implications for ESP rugby courses and provides recommendations for future research.

In summary, various methods have been employed to identify multiword items, with the most recent approach being a mixed-method corpus analysis and qualitative semantic analysis. The present study utilizes corpus-based and semantic approaches, adopting Nation's (2012) BNC/COCA base lists to compile a single-word list from the Masters' dissertations of Saudi engineering students. Subsequently, frequency, specialised occurrence, and semantic rating criteria were applied to identify multiword items within the corpus (see selection criteria in Chapter 3 for further details). The following section reviews several specialised studies focused on developing lists of technical vocabulary, including both single-word and multi-word items. This review will help identify a practical gap that the present study aims to address.

2.4.3.5 Engineering-Specific Word Lists

Building on the discussion of various 'discipline-specific' word lists, which include both single-word and multiword items across multiple disciplines, this section offers an overview of studies focused on engineering-specific word lists available in the literature. This review establishes the necessary background and highlights the gaps that the present study seeks to address. An overview of these studies is provided in Table 2.4 below.

Table 2-4: Engineering-Specific Word Lists

Author(s)	Word list	Corpus size	Text type (Genre)	Counting unit	Inclusion criteria	Size of list
Ward (1999)	The Engineering List (EngList)	1,000,000 words	Engineering course materials	Word family	Frequency	3,000-word families
Mudraya (2006)	Student Engineering Word List	Nearly 2,000,000	13 textbooks of basic engineering disciplines (BED)	Word family and word types	Frequency	1,200-word families or 9,000 word-types
Ward (2007)	Technical Collocations in Engineering Textbooks	Two corpora: A 380,000 words B 250,000 words	Corpus 1: Three chemical engineering textbooks Corpus 2: Textbooks from five engineering disciplines (chemical, civil, electrical, industrial and mechanical engineering)		Keyness, frequency analysis, and collocation analysis	78 noun phrase collocations
Ward (2009)	Basic Engineering	271,000 words	25 textbooks from 5 major	Word types	Frequency, specialised occurrence	299-word types

	English Word List		engineering sub-disciplines (chemical, civil, electrical, industrial, and mechanical)			
Jin, Ling, Tong, Sahiddan, Philip, Azmi & Tarmizi (2013)	Engineering Technology Word List (ETWL)	124,584 words	Engineering technology textbooks for Malaysian upper secondary vocational schools	Word types	Exclusion of GSL and AWL, specialised occurrence	313-word types
Hsu (2014a)	English Engineering Word List	4,570,000 words	A corpus of 100 engineering textbooks	Word family	Frequency and range (exclusion of GSL)	729-word families
Fox and Tigchelaar (2015)	Engineering Academic Formulas List	15 million running words	Published engineering research articles (RAs)	-	Keyness, frequency analysis	99 formulas
Watson Todd (2017)	An Opaque Engineering Word List	1.15 million tokens	Engineering textbooks from all 27 courses for undergraduate students	Lemma	Frequency, opacity	186-item keyword list
Coxhead, McLaughlin, and Reid (2018)	The Technical Word List (of Fabrication)	185,570 running words 99,000 running words (spoken) From 19 recordings totalling 26 hrs by fabrication tutors.	17 booklets from the Level 3 New Zealand Certificate in Mechanical Engineering	Word types	Frequency and semantic rating scale Nation's BNC/COCA (2012, 2016) frequency lists; frequency cut-off point (10 or more in the corpus); items from the fabrication-only file had to occur four or more times.	1,079-word types

Đurović. (2021)	Marine Engineering Word List	1,769,821	Instruction books and manuals of various ship and machinery types	Word family	Frequency and specialised occurrence	337-word families 73 transparent compounds

Table 2.4 provides a summary of research studies that have developed engineering-specific word lists. It includes details such as the author(s), the word lists created, corpus sizes, text types (genres), counting units, inclusion criteria, and the sizes of the resulting lists. To begin with, Ward (1999) developed the Engineering Corpus, a 1-million-word corpus from texts in first-year engineering courses at Suranaree University of Technology, Thailand, including thermodynamics, mechanics, and materials. This corpus contained 15,000-word types, grouped into word families at Level 6. A frequency analysis yielded the Engineering List (EngList), comprising the 3,000 most frequent word families. Validation showed significant differences between EngList and West's GSL (1953) and Xue & Nation's UWL (1984), with many EngList words not appearing in the GSL. Despite its technical focus and inclusion of essential general words, limitations included its reliance on non-specialised texts and the exclusion of other basic engineering sub-disciplines. Results indicated that first-year engineering students could achieve 95% text coverage with 2,000-word families.

Mudraya (2006) developed the Student Engineering English Corpus (SEEC), a 2-million-word corpus comprising 1,200-word families or 9,000-word types. The corpus was compiled from mandatory engineering textbooks across 13 disciplines, such as Engineering Mechanics, Engineering Materials, Mechanics of Materials, Fluid Mechanics, Thermodynamics, Electrical Engineering, Engineering Drawing, Manufacturing Processes, and Computer Programming, used at Walailak University in Thailand. The goal of the SEEC was to design a lexical syllabus aligned with the English teaching objectives for engineering students. The study employed WordSmith Tools (Scott, 1996) to identify 1,260 frequent-word families (appearing at least

100 times) in line with Bauer and Nation's (1993) Level 7. To validate the SEEC, comparisons were conducted with the COBUILD Bank of English Corpus and the BNC.

Additionally, Mudraya (2006) employed a data-driven approach, combining the lexical approach with corpus linguistics methodology, to enhance engineering students' language experience and increase their language awareness. To achieve this, the author incorporated sample data-driven instructional activities aligned with the lexical approach, aiming to help students acquire formulaic multi-word units or collocations, both for technical and non-technical purposes. The study emphasizes the importance of focusing on sub-technical vocabulary and academic English in ESP classrooms. However, unlike more recent studies in specialised disciplines, it did not explicitly create a dedicated multi-word list.

Ward (2007) developed an engineering multiword unit list called the Technical Collocations List to assist engineering students in transitioning from foundational topics (e.g., thermodynamics and structural analysis, among others) to more specialised fields (e.g., biomedical or aerospace engineering). The study investigated the lexical characteristics of undergraduate textbook content in chemical engineering and compared it with civil, electrical, industrial, and mechanical engineering. Two specialised corpora were constructed: Corpus 1, comprising approximately 380,000 tokens, was compiled from extensive extracts of three chemical engineering textbooks, while Corpus 2, containing around 250,000 words, was derived from five engineering disciplines (as specified). For Corpus 2, a random sample of 10,000 words was taken from each of the five textbooks in each discipline, resulting in a total of 25 samples of 10,000 words each. The chemical engineering sub-corpus included five textbooks covering eight of the thirteen courses offered to chemical engineering undergraduates at the author's university (some textbooks addressed multiple courses, such as Process Design 1 and Process Design 2). Using Scott's (2007) Wordsmith concordance tool, frequency and collocation analyses were conducted to identify multiword items in the two corpora. The study identified 78 noun phrase collocations, consisting of four four-word, 19 three-word, and 55 two-word collocations. However, the frequency cut-off point indicated that only those noun phrases that appear three times or more in the corpus were selected for inclusion in the list. Ward (2007) also compared the generated collocation list with other sub-disciplines and found that the Technical Collocations List was underused or seldom appeared in other sub-disciplines. This study highlighted that collocations can be specific to sub-disciplines. One limitation of the research is the relatively small size of the corpora analysed. Additionally, the study relied heavily on frequency-based analysis. Despite these constraints, the findings underscored the importance of technical vocabulary in developing subject-specific knowledge, extending beyond single-word units.

Ward (2009) developed the Basic Engineering List (BEL) to support less proficient foundation engineering undergraduates. The BEL was developed from an Engineering Corpus containing 250,000 running words, extracted from engineering textbooks in five engineering disciplines: chemical, civil, electrical, industrial, and mechanical. To ensure corpus representativeness, random pages (approximately 10,000 words each) were selected from 25 textbooks, covering the five engineering fields. The BEL was generated from the Engineering Corpus consisting of 271,000 running words selected from engineering course books recommended by five engineering faculties: chemical, civil, electrical, industrial, and mechanical. For corpus representativeness, random pages (approximately 10,000 words) were selected from each of the 25 textbooks. Ward used frequency and range as selection criteria, excluding function words. The BEL comprised 299-word types, accounting for 16.4% of the corpus. A notable limitation of the BEL is that it was developed ‘from scratch’, meaning it was not derived by excluding general words, despite being designed for weaker learners with limited vocabulary. Another distinguishing feature compared to earlier engineering word lists is that Ward used word types, rather than word families, as the unit of selection. This decision was based on the argument that different inflected forms can vary significantly from each other and from the headword. Ultimately, the BEL served as a practical resource for foundation-level engineering students who may not have been adequately prepared linguistically during high school to tackle English-language textbooks. Although the list is concise and non-technical, it offers broad coverage of engineering textbook content.

Jin et al. (2013) developed the Engineering Technology Word List (ETWL) to assist vocational secondary education in Malaysia. They compiled a specialised corpus of 124,584 words from engineering technology textbooks used in Malaysian upper secondary vocational schools. The ETWL comprises 313-word types, representing 8.7% of the corpus, and excludes general English function words, in addition to terms from West's (1953) GSL and Coxhead's AWL. This resource aims to assist in curriculum development for vocational engineering programmes in Malaysia.

Hsu (2014a) developed the English Engineering Word List (EEWL) using a corpus of 4.57 million running words, drawn from 100 college textbooks spanning 20 engineering disciplines, which were sourced from e-book databases. Unlike Jin et al. (2013), she only excludes West's (1953) most frequent 2,000 words and then applies strict range and frequency criteria. This study intended to examine the vocabulary demands of English-medium engineering textbooks and to develop an Engineering English Word List (EEWL) to address the lexical challenges faced by EFL students. The EEWL comprises 729-word families, accounting for 14.3% coverage of the textbook corpus from which it was derived. Similar to Ward's (1999, 2009) and Mudraya's (2006) lists, the target users of this list are undergraduate students; however, it is designed for more proficient learners who are assumed to already know high-frequency words (which were excluded from the list). After excluding West's (1953) GSL, the 729 most frequently occurring word families in the corpus were selected, achieving a 14.3% coverage threshold of the total words in the corpus. The study findings revealed that engineering undergraduate students can achieve 95% lexical coverage of an engineering textbook by mastering the most frequent 5,000-word families, along with proper nouns, transparent compounds, and abbreviations. Frequency analyses further reveal that students in civil engineering and mechanical engineering require knowledge of 3,500-word families for minimally acceptable comprehension, while those in marine engineering and biochemical engineering need up to 8,500-word families to reach the same level of understanding. Among the limitations, the EEWL missed some words with technical meanings that might appear in the GSL. It also uses word families instead of word types. These gaps will be addressed in the present study.

Fox and Tigchelaar (2015) developed a multiword items list, namely the Engineering Academic Formulas List, by partially replicating Ellis, Simpson-Vlach, and Maynard's (2008) research on the Academic Formulas List (AFL). The study utilized a corpus comprising 1 million running words extracted from published engineering research articles. For comparative analysis, a separate corpus of 1.5 million words from general academic discourse was also employed. A total of 99 formulas were identified based on the following criteria: (1) they were highly frequent, recurring expressions extracted from the 1-million-word engineering corpus; (2) they appeared significantly more often in the engineering corpus compared to the general academic corpus; and (3) they were found across a broad range of engineering subfields and publications. Initially, 765 formulas met the specified criteria (e.g., *in the presence of*). These were then evaluated by 12 graduate-level engineering teaching assistants, who rated them based on their pedagogical usefulness for newcomers to the engineering disciplinary discourse community. The evaluation used a Likert scale ranging from 1 (disagree) to 6 (agree). From this process, 99 formulas were ultimately selected. These were categorized according to their discursive functions: referential expressions e.g., *under, controlled,* and *conditions*, stance expressions e.g., *it is likely that*, and discourse-organizing expressions (Biber et al., 2004). A correlation analysis revealed connections between the formulas with the highest mutual information scores, their frequency in the corpus, and their perceived pedagogical value.

Watson Todd (2017) developed the Opaque Engineering Word List (OEWL) from a 1.15-million-word Engineering English Corpus (EEC), sourced from textbooks for 27 engineering courses in a Thai university. The list identifies 'opaque words' – polysemous terms whose engineering-specific meanings differ from their general usage. Using keyword frequency analysis with the BNC as a benchmark, 500 words were initially shortlisted, and filtered through five rigorous stages to yield 186 opaque words. These were rated for opacity using six criteria, including discrepancies in parts of speech and meanings between the EEC and reference dictionaries. Although the OEWL is small due to strict selection criteria, it is useful for teaching, focusing on challenging words for learners, but some key technical terms may be excluded.

Coxhead, McLaughlin, and Reid (2018) developed a subject-specific word list for one of the branches of engineering (fabrication). A fabrication corpus was created, containing 185,570 running words (written) and 99,000 running words (spoken) from 19 recordings totalling 26 hours by fabrication tutors. This word list was developed using word types as the unit of counting. Frequency and semantic rating served as the primary selection criteria. The authors adopted Nation's BNC/COCA (2012, 2016) frequency lists, setting a frequency cut-off point of 10 or more occurrences in the corpus. For items from the fabrication-only file, a lower threshold of four or more occurrences was applied, as these words have a narrower usage range. Experts were engaged to validate the word list using Chung and Nation's (2004) semantic rating scale. Ultimately, 1,079-word types were identified as part of the Fabrication Word List in the corpus.

The most recent engineering word list was developed by Durovic (2021). This is a subject-specific word list in the marine engineering discipline. This study's aim is two-fold: first, to develop a marine engineering word list, and second, to provide undergraduate marine engineering students with a practical vocabulary resource to help them achieve a 95% text coverage threshold for adequate reading comprehension. The author compiled a corpus of marine engineering instructional texts consisting of 1,769,821 running words. Nation's (2012) BNC/COCA frequency lists were used to assess the lexical load of the marine engineering corpus and establish the frequency cut-off point. The final list includes 337-word families and 73 compounds, collectively covering 8.13% of the target corpus. The study achieves 95% coverage by combining the 3,000 most frequent English words from the BNC/COCA lists with an expanded list of proper nouns, marginal words, and abbreviations. Like the approach planned in the present study, Durovic (2021) chose not to apply 'range' and 'dispersion' criteria, arguing that these measures are more relevant to large corpora and extracting general vocabulary across multiple professional fields rather than focusing on a single specialised domain.

Overall, seven different engineering word lists found in the literature were presented and discussed above. Some limitations were highlighted, especially regarding the principles of selection. It is also worth noting that these lists were created with different purposes and tailored to different target learner groups. For example, Jin et al.'s (2013) list was created for learners at the vocational secondary education level, while Ward's (1999, 2009) and Mudraya's (2006) lists are intended for undergraduate students. Watson Todd (2017) focuses on teachers and the highly technical frequent words that students might need to know. In terms of the text types (genres), all the corpora used in the previous engineering word lists were sourced from either course materials or textbooks. There is a lack of studies using other text types, such as research articles or dissertations, which are highly relevant forms of texts dealt with at the university level. As discussed in this section, frequency and specialised occurrence are the most adopted principles of selection, and range and dispersion were not recommended in this kind of research.

The present study employs a corpus-based methodology to develop a discipline-specific word list tailored to the field of engineering, with a particular focus on analysing masters' dissertations written by Saudi engineering students. This aims to develop a pedagogically useful resource for educators and students, especially those preparing to write their dissertations. By identifying technical vocabulary within engineering masters' dissertations, the study aims to create a valuable tool. Teachers play a crucial role in evaluating the list's usefulness and effectiveness, ensuring it meets classroom needs. Ultimately, this research contributes to vocabulary studies by advancing the development of an engineering-specific word list, offering practical tools to enhance language learning and teaching in engineering education.

2.5 A Disciplinary Approach to Written Discourse

This section presents a discussion about a disciplinary approach to a written discourse, which is an important aspect in the development of this thesis, with a focus on engineering masters' dissertations as a specialised genre in academic discourse. The terms *discipline*, *disciplines*, and *disciplinary* are terminologies that are 'discoursally constructed' (Thompson & Hunston, 2020). For instance, the term 'discipline' can refer to 'a field of study' or "the body of knowledge linked to a specific field of study" (Holley, 2009). In the context of this study, the concept of ‘a discipline as a discoursally constructed entity’ is particularly significant, emphasizing the academic exploration of disciplinary language.

For laypeople or novices entering a specific professional field, such as engineering, to effectively engage in a discipline-specific discourse community, they must first be introduced to and acquire the language commonly used within that community (Flowerdew, 2002; Hyland, 2015). Recent research has consistently demonstrated that academic discourses are tremendously diverse, and this diversity has crucial implications for writers as they interact with their teachers and colleagues, in addition to writing themselves into their subjects (Hyland, 2004). Bhatia (2002) emphasizes that for students to master discipline-specific written conventions, they must first develop an understanding of the discursive processes and practices. This foundational awareness enables them to connect with the structures, purposes, and social environments of their future professional communities (Tardy, 2009). Such awareness can be cultivated through explicit vocabulary instruction (Fox & Tigchelaar, 2015; Li & Schmitt, 2009; Tardy, 2009).

Research into discipline-specific writing is increasingly providing valuable insights into the forms and functions of written texts across various fields, in addition to how different academic contexts impose demands on communicative practices that are often unfamiliar to disciplinary newcomers (Hyland, 2004). This research is instrumental in understanding how disciplines are continually reshaped through individuals’ ongoing social and discursive interactions as they negotiate and reaffirm their membership within academic communities. Effective academic writing depends on the writer’s capacity to convey a sense of shared professional identity. In other words, writers seek to situate their work within a specific social framework, which they evoke and reflect through the use of discipline-specific, accepted discourses to achieve their personal and professional goals (Hyland, 2004). Academic written genres have garnered significant attention from fields such as philosophy, science, sociology, history, rhetoric, and applied linguistics. Despite this, there is widespread consensus on the central role of written texts in academic life, with the recognition that understanding a discipline requires a deep comprehension of its discourses for two primary reasons (Hyland, 2004).

One of the reasons is that disciplinary discourse served as a crucial source of insight into the social practices of academic communities. For instance, discourses are systematically structured sets of statements that articulate the meanings, values, and norms of an institution (Hyland, 2004). Thompson and Hunston (2020) highlight that research on disciplinary variation has garnered significant attention over the years, with the majority of studies being conducted within the field of EAP. This focus underscores the importance of understanding how language functions within specific academic contexts to facilitate effective communication and knowledge construction.

Writing, in fact, draws on all of the professional resources, navigating through the various constraints, structures, and dynamics that characterize the professional domain particular to the profession. Hyland (2008) argues that research articles serve as a key resource for identifying the most frequently used language within a discipline, and this can be effectively achieved through corpus analysis. Researchers, particularly those adopting corpus approaches at the discoursal or genre level, are chiefly concerned with patterning, rhetorical moves, stance, and phraseology in academic writing (Flowerdew, 2002). Some researchers have focused on comparing L1 and L2 academic writing. For instance, Nasseri and Thompson (2021) explore lexical density and diversity in dissertation abstracts, while Stamatović, et al., (2020) analyse the vocabulary used in graduation theses written by English philology students, comparing the academic writing of Montenegrin and US students. In contrast, Thompson (2005) investigates the use of intertextual references in eight PhD theses from the Agricultural Botany department at a British university. Meanwhile, Therova (2021) concentrates solely on L1 students in her study, titled A Corpus-Based Study of Academic Vocabulary in Foundation-Level Students' Assessed Academic Writing at a UK University. Hence, the present study does not compare L1 and L2; rather, it focuses on academic writing by EFL students from Saudi Arabia. A corpus of masters' dissertations was developed (as a specialised academic genre). The following section discusses corpus-based genre approaches to academic writing.

2.5.1 Corpus-Based Lexical Analysis of Written Academic Genres

According to Hyland (2005), the term '*genre*' refers to the classification of texts that reflect how writers typically use language to address recurring situations. Swales' (1990) seminal work on genre analysis, which explores core concepts such as genre, discourse community, and task, is widely regarded as foundational in the field (Chang & Kuo, 2011). The concept is rooted in the idea that members of a community can easily recognize similarities in the texts they frequently encounter, enabling them to read, understand, and often produce such texts with relative ease due to their repeated exposure. Genres are thus understood as "the outcomes of individual social agents acting within the boundaries of their historical and contextual constraints while drawing on their knowledge of existing generic forms" (Kress, 1989, p. 10). Texts are crafted to be interpreted within specific cultural contexts, and genre analysis can reveal the underlying norms of these academic cultures, with conventional rhetorical practices reflecting individual writers' perceptions of shared values (Hyland, 2010). Consequently, genre theorists emphasize the centrality of participants' interactions in language use, assuming that an effective text will demonstrate the writer's awareness of its context and the audience within that context (Hyland, 2005).

The work by Biber (1988) has had a major influence on the corpus linguistics method regarding genre study, which was further expanded by Biber and Conrad (2009), who used two major approaches – genre and register – to text analysis. A genre approach, as defined by Biber and Conrad (2009), involves characterizing text varieties based on their rhetorical organization. This implies that linguistic analysis of texts from a genre perspective centres on the conventional frameworks used to create complete texts. Consequently, the features of a genre are considered conventional rather than merely common. As a result, genre studies in this field emphasize detailed and comprehensive text analysis (Egbert & Gray, 2019). Complete texts are essential for analysing linguistic characteristics from a genre perspective, which emphasizes the structural features used to organize entire texts. In contrast, text excerpts are sufficient for register analysis, which focuses on identifying frequent and pervasive linguistic features distributed throughout the text (Egbert & Gray, 2019). It is important to note that full texts examined from a genre perspective can also be analysed from a register perspective (Biber, 2019). However, the issue of whether to include complete texts or some parts of them depends on the nature of the researcher's question and the ethical issues addressed in conducting a particular study. For example, Bunton (1998) explored the rhetorical moves within the introduction and conclusion sections of PhD theses. The present study includes the complete texts of the theses, excluding acknowledgements and appendices.

Several studies on university student writing have employed corpus linguistics methodologies by creating learner corpora – collections of texts produced by learners. These corpora include language data from learners of different ages, linguistic backgrounds, proficiency levels, and regions, enabling the exploration of a wide range of research topics (Díaz-Negrillo, Thompson & Ballier, 2013). For example, Staples and Reppen (2016) compiled a corpus of approximately 400,000 words of first-year university student writing across three first languages (English, Chinese, Arabic), two genres (rhetorical analysis and extended argument), and various language proficiency levels, with a focus on vocabulary and grammar usage. Their findings reveal similarities in lexico-grammatical features among L1 writers.

Different academic genres have been involved in corpus-based lexical studies, such as students' academic essays (e.g., Hinkel, 2004; Johns, 2002; Hyland, 2004) and research articles as distinctive academic written genres (e.g., Hyland, 2000). However, in terms of length, research articles tend to be shorter types of academic writing with relatively few sections compared to theses. There are relatively few studies employing a genre-based approach to master dissertations and PhD theses (Charles, 2006; Loghmani et al., 2020; Xiao & Sun, 2020; Thompson, 2005), which are discussed in detail in the next section as they align closely with the focus of the present research. This present study specifically aims to investigate the technical vocabulary in engineering masters' dissertations written by Saudi students.

2.5.2 Dissertation/Thesis as a Distinctive Genre of Academic Writing

Recently, there has been an increasing focus on research examining academic texts, particularly masters' dissertations and doctoral theses, within the field of EAP (Paltridge & Starfield, 2020; Sun & Crosthwaite, 2022; Thompson, 2005). However, much of this research has focused on meta-discourse markers, such as hedges, in these academic genres (Charles, 2006; Loghmani, Ghonsooly & Ghazanfari, 2020; Xiao & Sun, 2020). Few studies have investigated the overall discourse structure of dissertations and theses, with most research concentrating on specific sections. Notable exceptions include Dudley-Evans (1999) and Thompson (1999), who analysed the structure of these texts in their entirety. However, none of these studies specifically examine the lexical profile of vocabulary.

Several studies have examined the linguistic, stylistic, and structural characteristics of academic abstracts as a distinct sub-genre (e.g., Gillaerts & Van de Velde, 2010; Lee & Casal, 2014; Nasseri & Thompson, 2021). Nasseri and Thompson (2021) highlight that the abstract section of a thesis, dissertation, or research article is widely regarded as a unique sub-genre of academic writing in numerous significant studies. An abstract is characterized as a lexically dense summary that provides an overview of the entire thesis or article, briefly outlining the introduction, methodology, key findings, and the broader impact, applications, and contributions of the study and its outcomes (Bitchener, 2009; Gillaerts & Van de Velde, 2010).

Lee and Casal (2014) conducted an analysis of a corpus comprising 200 results and discussion chapters from masters' dissertations in engineering, authored by English and Spanish writers. Their findings emphasize that the interpersonal aspects of academic writing are shaped by the specific linguistic and cultural contexts in which the texts are produced, even within the same field. Similarly, Sun and Crosthwaite (2022) explored discipline-specific forms and functions of negation in the 'limitations' sections of 120 PhD theses. Utilizing the appraisal framework, their study covered disciplines classified as hard-applied, hard-pure, soft-applied, and soft-pure, uncovering variations in the use of negation across these categories.

Alharbi (2021) examined meta-discourse features in Applied Linguistics academic research articles and master dissertations, applying Hyland's (2005a) meta-discourse model to analyse the identified elements. The study's findings demonstrated that interactive meta-discourse features were more prevalent than interactional features in both text types. However, the master dissertations sub-corpus showed significantly higher occurrences of most meta-discourse devices compared to research articles. Among these, transitions emerged as the most commonly used category of meta-discourse, while hedges stood out as the most common interactional meta-discourse feature in both sub-corpora.

Hyland and Tse (2005) investigated the frequency and usage of '*evaluative that*' in a corpus of abstracts from research articles, master dissertations, and PhD theses written by English as a Second Language authors, exploring its various forms and functions. Other studies have analysed abstracts from research articles and conference papers, focusing on elements such as tense, voice, modal and reporting verbs, stance words, nouns, *that*-complement clauses, and other linguistic features. However, the abstract section of master dissertations remains an underexplored area, particularly in terms of differences in lexical richness and diversity between abstracts written by English L1 and L2 students. Additionally, the constraints imposed by word limits, which require abstract writers to convey key concepts concisely while limiting the use of non-repetitive vocabulary, further complicate this issue (Hyland & Tse, 2005).

Taymaz (2021) employed a corpus-based methodology to explore shifts in the frequency of hedges and boosters in the academic writing of Turkish EFL students as they advanced from master to PhD levels within the field of English Language Teaching. The study examined the discussion sections of ten master dissertations and ten PhD theses, all authored by the same students, to assess potential developmental changes between the two academic levels. The findings revealed a higher frequency of boosters in masters' dissertations compared to PhD theses, while hedges were more frequently used in the latter. These results suggested that the increased use of boosters at the PhD level may reflect greater self-confidence in conveying ideas, supported by broader academic experience and expertise. Conversely, the higher use of hedges at the master level might indicate less certainty. The findings also highlighted a degree of unawareness among both master and PhD students regarding the usage of meta-discourse markers and academic writing conventions.

Although a significant amount of corpus-based research on thesis writing has examined diverse linguistic features across various sections, there is still a noticeable lack of studies focusing on the lexical profiles and vocabulary load of master dissertations and PhD theses, especially in relation to technical vocabulary. The use of technical vocabulary reflects the mastery required in specialised disciplines, such as engineering, which is the primary focus of this thesis. Paltridge and Starfield (2020) suggested that thesis writers must be proficient in their field of specialization to gain acceptance into their respective academic communities. Therefore, the present thesis intends to investigate vocabulary used in engineering masters' dissertations as a specialised genre. It also examines the distribution of technical vocabulary across the sections of these dissertations.

2.6 Teachers' Perceptions of Vocabulary Learning

Teachers have a significant influence on how vocabulary is taught and learned, and they can offer useful information for word list evaluation (Benson, 2020). Engaging teachers in assessing vocabulary in EAP learning materials would yield valuable insights into how vocabulary is considered during the design and adaptation of materials for EAP learners (Lu & Dang, 2022). Teachers' classroom practices are shaped by their perceptions or cognition (that is, their thoughts, beliefs, and knowledge), which are influenced by numerous factors. Despite the importance of teacher cognition, studies examining teachers' perspectives on vocabulary learning and teaching are very limited in number (Dang & Webb, 2020; He & Godfroid, 2018; Lu & Dang, 2022).

Lu and Dang (2022) emphasized that insights from both students and teachers can offer valuable perspectives in text analysis, enriching the findings of quantitative corpus-based research through the inclusion of qualitative data. In their study, Lu and Dang (2022) used semi-structured interviews with six EAP instructors, revealing that many teachers found the terminology in learning materials to be overly challenging for students. This challenge was linked to an excessive vocabulary load, which may have arisen from inadequate attention to vocabulary selection during material development (Coxhead, 2018). The findings suggest that this issue may stem from a lack of experience or limited familiarity with research-based principles of vocabulary instruction, which could undermine the pedagogical effectiveness of these materials.

Dang, et al., (2022a) investigated the perceived usefulness of high-frequency English vocabulary from the perspectives of both teachers and learners, emphasizing the critical role teachers play in evaluating word lists. Their findings revealed that teachers considered the BNC/COCA2000 word list more useful, and learners showed greater familiarity with its vocabulary. This indicates that the BNC/COCA2000 is the most effective high-frequency word list for L2 learners.

In a related study, Dang, et al., (2022b) examined the perceptions of 78 EFL teachers regarding the usefulness of 973 high-frequency words. They analysed correlations between teacher perceptions and factors such as lexical coverage and learners' receptive vocabulary knowledge, identifying strong associations. Importantly, teacher perceptions were found to be a strong predictor of learners' knowledge of high-frequency vocabulary. These results highlight the significant influence of teachers' judgements on vocabulary acquisition in EFL contexts. Therefore, the present study aims to explore engineering teachers' perceptions of the usefulness of ETV, with the goal of developing a comprehensive list for pedagogical application.

2.7 Summary of the Chapter

This chapter reviewed the literature relevant to the research questions of this study, which aimed to investigate the knowledge and use of technical vocabulary in masters' dissertations written by Saudi engineering students. Specifically, the review focused on key areas related to technical vocabulary, including its identification, classification, and pedagogical applications, while identifying gaps that the present study seeks to address.

The chapter explored aspects of technical vocabulary, particularly the challenges in profiling technical vocabulary, such as identifying single-word and MWUs, determining units of counting, and establishing selection criteria. Since one of the objectives of this study is to identify both single-word and MWU engineering technical vocabulary, the chapter discussed criteria used in previous specialised studies and highlighted gaps in existing methodologies that this research aims to address.

Additionally, the chapter reviewed various types of word lists developed in vocabulary research, identifying limitations in their applicability to engineering contexts. It also examined studies on teacher perceptions of vocabulary usefulness, emphasizing the need for greater involvement of teachers in evaluating the pedagogical value of technical vocabulary. This gap is addressed in the present study by incorporating teacher evaluations of the usefulness of engineering technical vocabulary.

Finally, the chapter concluded by summarizing how this research will contribute to filling these gaps. The next chapter provides an overview of the research methodology employed to address the study's research questions.

Chapter 3 : Methodology

3.1 Introduction

This empirical research aims to investigate the knowledge and the use of technical vocabulary in a corpus of Saudi engineering students' Masters' dissertations. The study consists of three linked studies which were conducted to address the following objectives. First, to find out more about the nature of vocabulary engineering discipline in a specialised genre of Saudi masters' dissertations, study one entails a corpus-based analysis of a lexical profile and vocabulary load analysis of the EMDC. Building on study one, the second study focuses on the identification of ETV in EMDC in terms of both single-word items and multiword units. Finally, in the third study, The ETV items identified in Study 2 were used to investigate the students' receptive knowledge of engineering vocabulary, exploring also the perspective of engineering teachers in terms of their perceptions of the pedagogical usefulness of the ETV list.

3.2 Building the Engineering Masters' Dissertations Corpus (EMDC)

Building a representative corpus that can accurately reflect linguistic variety and appropriately represent language use in the target discourse area is a crucial aspect of corpus-based research. This section outlines the procedures for building the EMDC.

3.2.1 Constructing a Representative Corpus

Constructing a representative corpus that reflects authentic language use and linguistic variation in a specific discourse domain is one of the most challenging aspects of corpus-based research (Sinclair, 1991). When constructing a corpus, size and representativeness are crucial issues that can influence the research's validity and reliability (Biber 1993; Sinclair, 1991). Therefore, the size and representativeness of the corpus were taken into account when compiling the EMDC, as suggested by Nation (2016) regarding the development of vocabulary lists.

Concerning size, a smaller corpus is inherently less reliable (Sinclair, 1991). As Flowerdew (1996) suggests, the larger the corpus, the better it is for extracting linguistic information. Thus, a corpus that consists of larger numbers of texts is more likely to give an accurate representation of the language sample. However, determining an appropriate corpus size largely depends on the nature of the research, such as whether the corpus is spoken or written. Spoken corpora typically contain fewer running words than written corpora, as noted in Benson and Coxhead (2022). Nevertheless, there are no defined rules for establishing/setting, the total size of any corpus for a specific purpose, particularly for terminology extraction. Kennedy (1998) suggests that a corpus for specialised purposes should contain at least 100,000 running words to allow researchers to explore the statistical (probabilistic) nature of lexical elements within the text. Conner and Upton (2004) propose that a range of 20,000 to 250,000 words can be regarded as a sub-corpus or small corpus. However, more recent research on word lists tends to favour larger corpora to ensure greater representativeness and more in-depth analysis (e.g., Brezina & Gablasova, 2015; Gardner & Davies, 2014). Regarding the size of the target corpus, it contains over one million running words, as presented in Table 3.1 (Overview of the Cleaned Cleaned EMDC).

The three principles for corpus design proposed by Biber (1993) were followed in order to achieve representativeness: (1) defining the target population that the corpus is intended to represent; (2) choosing a sampling framework to direct the selection of texts; and (3) assessing the degree to which the sampled texts reflect the target population.

3.2.1.1 Target Population

The target population for the EMDC consists of a collection of masters' dissertations written by Saudi students in engineering programmes at King Abdulaziz University, dated to 2022. KAU's Faculty of Engineering, established in 1969, offers a wide range of English-medium master’s programmes, including but not limited to Aeronautical, Chemical, Civil, Electrical, Industrial, Mechanical, and Nuclear Engineering. The completion of a dissertation is a mandatory requirement for degree conferral in these programmes. This body of dissertations constitutes the population from which the EMDC was sampled. Having defined the target population, the next step was to determine an appropriate sampling strategy to build a representative and manageable corpus.

3.2.1.2 Sampling Technique

The sampling strategy was guided by the principles of specialised corpus design, primarily following the framework established by Biber (1993). The process involved three key stages. First, a sampling frame was established. KAU was selected as the single source institution to ensure disciplinary and institutional homogeneity, a crucial factor for creating a specialised corpus. The sampling frame comprised all electronically available masters’ dissertations from the university’s library archive across all engineering disciplines.

Second, a principle of Biber’s framework is determining the size of the target corpus. In this study, the target size of the EMDC was guided by recommendations from Corner and Upton (2004), suggesting a range between one and five million words for specialised corpora. This range is deemed sufficient for the reliable identification of core technical vocabulary (Lu & Durrant, 2017; Todd, 2017). The EMDC consists of 1,322,437 words (see Table 3.1).

Third, a practical sampling procedure was implemented. All accessible dissertations within the sampling frame were retrieved as PDF files. This comprehensive retrieval approach aimed to maximize the corpus size and representativeness within practical constraints. The files were then converted to plain text using Anthony’s (2022) AntFileConverter tool and underwent a meticulous manual cleaning process to remove formatting artifacts, tables, figures, and front/back matter, isolating the core analytical text for analysis. The final cleaned texts were compiled to form the EMDC.

However, reflecting on the sampling methodology, two primary limitations are acknowledged. First, the corpus may not be fully representative of the entire target population due to variable institutional maturity across engineering sub-disciplines. Some specializations (e.g., Nuclear Engineering) are newer at KAU, resulting in a smaller number of available dissertations compared to established fields (e.g., Civil or Electrical Engineering). Consequently, the vocabulary of these newer disciplines may be underrepresented in the EMDC, potentially limiting the generalizability of the resulting word list across all engineering sub-fields. Second, the sampling method was one of comprehensive convenience within the defined frame. While every effort was made to include all accessible texts, the final corpus is dependent on the availability and completeness of the university's digital archive, which may not contain every dissertation ever produced.

Despite these limitations, the corpus maintains a high degree of internal homogeneity, a key strength for a specialised corpus. It exclusively comprises dissertations written by a defined population (Saudi master's students) within a single institutional context (KAU) for the same academic purpose (fulfilling degree requirements). This homogeneity ensures that the analyzed lexicon is specific to the context of Saudi engineering academia, which is the central focus of this study.

3.2.2 Overview of the Engineering Masters' Dissertations Corpus

For the purposes of the present research, a target corpus was constructed through the collection of masters' dissertations written by Saudi engineering students at KAU. In terms of content, all dissertations included were master's level texts from seven engineering courses, showing the homogeneity of the data and its suitability to represent the engineering genre. These dissertations were submitted by students as partial fulfilment of the requirements for their master's programmes at the university. Each dissertation was structured into five chapters, introduction, literature reviews, methodology, results and discussion, and conclusion. The length of these dissertations varied, typically around 30,000 words, reflecting the nature of the engineering discipline, which often includes information presented in graphs, tables, and statistical data. These non-textual elements were excluded during the corpus construction process, as discussed in the next sub-section, corpus cleaning.

This specialised corpus created was named the Engineering Masters' Dissertations Corpus (EMDC). The EMDC data was generated from samples of different engineering fields offered at KAU, including Aeronautical Engineering, Chemical Engineering, Civil Engineering, Electrical Engineering, Industrial Engineering, Mechanical Engineering, and Nuclear Engineering.

The full texts of all dissertations were available in electronic format, as recommended by Sinclair (1991) and Barnbrook (2019). All dissertations were retrieved as PDF files and then converted to plain text using Anthony's (2022) AntFileConverter tool. Following the conversion process, the data underwent manual cleaning, before being analysed using the AntWordProfiler tool, as explained in the following section.

3.2.3 Corpus Cleaning

Text cleaning is an essential prerequisite for analysis, as it ensures the standardization of the corpus (Benson & Coxhead, 2022; Chen & Ge, 2007; Lu & Coxhead, 2020). The first cause for concern here pertains to excluding all bibliographies, acknowledgements, tables of contents, figures, formula indices, references, and appendices from the text files. These different elements are purposely removed to ensure the accuracy of the frequency analysis conducted in this research (McEnery & Brookes, 2024). For example, references were removed because they did not reflect the language production of the dissertation students. Figures and formula indices were not used because they are incompatible with textual analysis tools such as AntWordProfiler and Range.

Another consideration that arises during corpus cleaning relates to the correction of typos and spelling mistakes e.g., *lough*, *perliminaty*, *serics* (McEnery and Brookes, 2024; Nation, 2016) using a grammar checker. Additionally, contractions were checked during the process of cleaning the corpus. As suggested by Coxhead and Demecheleer (2018) and Benson and Coxhead (2022), contractions were reverted to their original condition after the analysis to preserve the corpus typical of the language in use. All the contraction forms were rewritten in full. For example, '*don't*' was rewritten as '*do not'*. Following these steps, the EMDC was created, containing a total of 1,322,437 running words, as shown in Table 3.1.

Table 3-1:Overview of the Cleaned Engineering Masters' Dissertations Corpus

Sub-fields of Engineering	Number of Dissertations	Number of Tokens
Aeronautical Engineering	6	68,163
Chemical Engineering	5	37,638
Civil Engineering	2	27,740
Electrical Engineering	43	455,982
Industrial Engineering	28	459,816
Mechanical Engineering	21	178,588
Nuclear Engineering	8	94,510
Total	113	1,322,437

As shown in Table 3.1, the cleaned EMDC data contained 1,322,437, which is large enough to capture the most frequently used technical vocabulary. Previous studies investigating technical vocabulary have developed target corpora of different sizes similar to the data size of the current study. For example, Lu and Durrant (2017) built a 1,045,969 Traditional Chinese Medicine Corpus from journal articles, Ward (1999) developed a 1-million-word corpus from engineering course materials, and Watson Todd (2017) created the Opaque Engineering Word List from 1.15 million words extracted from engineering textbooks. Therefore, the target size of the corpus in the present research is justifiable. It has been established that a corpus of one million words is substantial for word list generation (Brysbaert & New, 2009, cited in Qi, 2016).

3.2.4 Piloting the Engineering Masters' Dissertations Corpus

This section presents the pilot study, which describes the sample data analysis process to ensure that the EMDC and the adopted BNC/COCA could be processed appropriately by AntWordProfiler. Also, the pilot study helped to identified areas requiring further cleaning to enhance the accuracy and consistency of the data. To this end, ten masters' dissertations in the engineering discipline were taken from the main corpus to develop a mini corpus. The pilot corpus was processed using the AntWordProfiler tool, which facilitated the analytical process for the main study. One key issue identified in the sampled data during the pilot test was the need for further cleaning of the data for the main study, especially in terms of spelling. For example, the words *Arabia* and *reduces* appeared in some parts of the EMDC as *a rabia* and *reduc ces*, which would have increased the number of 'Not in the list' items.

3.3 Overview of Methodology in Study 1: Lexical Profiling of Vocabulary in the Engineering Masters' Dissertations Corpus

This section provides an overview of the methodology employed in Study 1, the main objective of which was to examine the lexical profile and vocabulary load of the Engineering masters' Dissertations Corpus using a corpus-based approach. The study aimed to identify general lexical features within this specialised genre. Study 1 addresses the following two research questions:

1. What is the lexical profile of the EMDC across Nation's (2012) BNC/COCA word lists?

   a. What is the coverage of Schmitt and Schmitt's (2014) high-, mid-, and low-frequency vocabulary in the EMDC?

2. What is the vocabulary load of the EMDC across Nation's (2012) BNC/COCA word lists?

To explore the characteristics of vocabulary in engineering masters' dissertations, Study 1 began with a lexical profile analysis of the cleaned EMDC using the AntWordProfiler tool. AntWordProfiler is one of the most up-to-date and widely recommended software tools developed specifically designed for conducting lexical profile analysis (Đurović & Vuković-Stamatović, 2021). Developed by Laurence Anthony (2014, version 14), it serves as an upgraded version of the previously used Range programmes (Nation & Heatley, 1994). In this study, the EMDC was entered into the programmes and its coverage was analysed according to the lexical coverage in Nation's (2012) BNC/COCA base word lists. These lists consist of word families ranging from 1,000 to 25,000 base word list levels, along with four supplementary lists (proper nouns, marginal words, transparent compounds, and abbreviations). To determine the coverage provided, the first step involved ranking each headword in the EMDC is ranked according to its frequency in the BNC/COCA lists.

The profile analysis was also used to examine the coverage of Schmitt and Schmitt's (2014) high-, mid-, and low-frequency vocabulary bands to gain deeper insights into the nature of vocabulary in the EMDC. Schmitt and Schmitt (2014) categorized Nation's (2012) frequency-based base word lists into three bands: high-frequency vocabulary (the first three 1,000-word families), mid-frequency vocabulary (the 4,000–8,000-word bands), and low-frequency vocabulary (the 9,000–25,000-word bands). The extent of coverage for each frequency band was calculated by adding the coverage percentages of the BNC/COCA base word lists that fall within that particular frequency band.

Furthermore, the profile analysis aimed to identify engineering-specific words from the words not present in Nation's (2012) BNC/COCA base word lists (1 to 25) or supplementary lists, following the methodology used in previous studies (e.g., Benson, 2020; Coxhead & Demecheleer, 2018; Lu & Dang, 2022; Lu, 2018). The process involved examining the vocabulary outside the existing BNC/COCA frequency-level lists or supplementary lists to identify potentially specialised vocabulary (Benson, & Coxhead, 2022). The researcher validated the 'off-list' words identified in the EMDC's preliminary profile analysis by consulting reference sources, including the *Merriam-Webster Dictionary* and *McGraw-Hill's Dictionary of Engineering* (2003). Words with specialised engineering meanings in the dictionary were subsequently added to the Engineering Technical Vocabulary list. Through this process, an additional base list named the ‘Engineering Base List’, was created to conduct a profile analysis and determine the lexical profile and vocabulary load using refined base word lists (See Chapter 4, Section 4.3.2.6 for details).

Finally, the study examines the vocabulary load of the EMDC, which refers to the amount of vocabulary needed for adequate reading comprehension (Đurović, 2021). Vocabulary load is considered the range of vocabulary essential for achieving optimal reading comprehension (Schmitt et al., 2011). The study adopts 95% and 98% coverage as the recommended vocabulary thresholds for minimal and optimal reading comprehension, respectively, as suggested by previous research (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Laufer, 1989; Van Zeeland & Schmitt, 2013). The lexical load analysis identifies the vocabulary needs of engineering learners to comprehend the EMDC without vocabulary being a handicap. Detailed procedures for data collection and analysis in Study 1 are presented in Chapter 4.

3.4 Overview of Methodology in Study 2: Identification of Technical Vocabulary in the EMDC

This section provides an overview of the methodology employed in Study 2. This study primarily focuses on identifying single-word and multi-word technical vocabulary in the EMDC. Additionally, it examines the distribution of technical vocabulary across the five sections of the masters' dissertations: Introduction, Literature Review, Methodology, Results and Discussion, and Conclusion.

Study 2 aims to address the following two research questions:

1. Which technical vocabulary items, in terms of single-word units, are used in the EMDC?
   a. How many of these belong to the high-, mid-, and low-frequency vocabulary bands, respectively?
2. Which technical vocabulary items, in terms of multiword units, are used in the EMDC?
3. What is the distribution of the engineering technical vocabulary list across the five different sections (Introduction, Literature Review, Methodology, Results and Discussion, and Conclusion) of the engineering masters' dissertations?

A key challenge in identifying technical vocabulary lies in determining whether a word is used with a general, everyday meaning or a specialised, technical meaning specific to the discipline (Coxhead et al., 2016). Past studies have shown that researchers have employed five different methods for identifying technical vocabulary from specialised corpora. These methods include consulting dictionaries (Nation, 2001; Oh, Lee, Lee, & Choi, 2000), using context clues (Flowerdew, 1992), conducting keyword analysis (Anthony, 2004; Scott, 1997), applying semantic rating scales (Chung & Nation, 2003), and utilizing hybrid methods (Kwary, 2011).

These five methods can be divided into two broad categories: corpus-based and judgement-based approaches (Liu & Lei, 2020).

3.4.1 Identification of Single-Word-Unit Technical Vocabulary in the EMDC

In this study, a mixed-methods approach was employed that combines corpus-based and judgement-based methods (Liu & Lei, 2020) to identify technical vocabulary from the compiled specialised corpus of engineering masters' dissertations. The identification of single-word technical vocabulary involved four key procedures: conducting keyword analysis, applying frequency principles (using frequency cut-off points), consulting the *McGraw-Hill Dictionary of Engineering* (2003), and analysing concordance lines when necessary. Following these steps, a judgement-based approach was employed whereby experts in the engineering domain rated potential technical vocabulary using Chung and Nation's (2003) semantic rating scales to evaluate whether the identified words qualified as technical vocabulary within the field.

Another method adopts is the identification of keywords in the EMDC, which was performed using WordSmith Tools 6.0 (Scott, 2012). In this Keyness analysis, the specialised EMDC served as the target corpus while the British National Corpus (BNC Written Corpus) functioned as the reference corpus representing general English usage. The selection BNC as reference corpus was deliberate and methodologically significant. According to Scott (1997), keywords are words that, when compared to a reference corpus, occur in a target corpus with a statistically substantial frequency. Using the general English BNC instead of another specialised corpus enhances the identification of genuinely specialised engineering terminology by maximizing the contrast between everyday language and discipline-specific usage. This methodological choice follows Scott's (2006) observation that technical terms typically exhibit markedly higher frequency in specialised domains than in general language.

This method has been commonly used in technical previous vocabulary studies (Benson & Coxhead, 2022; Chung, 2003; Coxhead et al., 2016; Coxhead & Demecheleer, 2018; Ha & Hyland, 2017; Quero, 2015; Watson Todd, 2017). Additionally, word type was chosen as the unit of identification for technical vocabulary because it is widely used in such studies (Benson & Coxhead, 2022; Chung, 2003; Coxhead et al., 2016; Coxhead & Demecheleer, 2018; Ha & Hyland, 2017; Quero, 2015; Watson Todd, 2017). Using word types rather than word families is recommended because individual word types may be technical while other members of the same word family may not be (Benson & Coxhead, 2022; Chung & Nation, 2003).

Thus, a list of all possible technical vocabulary items in the EMDC was generated using the keywords function of WordSmith Tools 6.0 (Scott, 2012), ordered according to their relative *Keyness* compared with the reference corpus (BNC Written Corpus). The focus here was on the keywords with a positive Log-Likelihood Ratio (LLR), as they were statistically more probable to occur in engineering discourse. The analysis output found 2,470 keywords with a positive LLR, with the ten most frequent positive words being the following: *figure, antenna, system, temperature, data, model, simulation, results,* and *energy* (see detailed analysis in Section 5.2.1).

The second approach, the frequency principle, was applied to select initial possible technical vocabulary from the keyword results. This principle entails a process of adopting frequency cut-off points "which allowed the maximum number of potential technical words to be identified" (Benson, & Coxhead, 2022, p.115). Frequency is the most prevalent criterion used for assessing whether a word is worthy of inclusion in a word list (Benson & Coxhead, 2022; Coxhead et al., 2017; Gardner & Davies, 2014; Yang, 2015). However, there is no universal frequency threshold that can be used in all corpus-based word list studies. The frequency threshold depends on the goal of the word list in each investigation, considering the cut-off points used by Coxhead (2000) with 100 occurrences in a corpus of 3.5 million words, Ha (2015) with 63 occurrences in a corpus of 6.7 million running words, and Lu (2018) with 60 occurrences in a corpus of 5 million running words. A cut-off point of 60 occurrences in the EMDC was then used to select potential ETV items from the positive keywords list for pedagogical purposes, resulting in the identification of 1,146 positive keywords out of the 2,470 keywords with a positive likelihood ratio identified using keyword analysis.

Additionally, dictionaries were consulted to check the meaning of all 1,146 positive keywords using the *McGraw-Hill Dictionary of Engineering* (2003) or other dictionaries, such as *the Merriam-Webster Online Dictionary* and *the Macmillan Dictionary*. Concordance lines were reviewed to analyse the functions of these words within context when necessary. If a word was found as a headword in the dictionary, it was directly categorized as technical vocabulary and added to the final list. For example, the word *'inverter'*, was found in the Engineering dictionary and defined as 'a device for converting direct current into alternating current; it may be electromechanical, as in a vibrator or synchronous inverter, or electronic, as in a thyratron inverter circuit' (McGraw-Hill Dictionary of Engineering, 2003).This step helped reduce subjectivity by relying on established engineering references to confirm the technical status of words, rather than subjective judgement by raters. Out of 1,146 positive keywords identified using the frequency principle, 536 words were found in the engineering dictionary. The remaining 613 words were then subjected to qualitative semantic analysis by independent raters who were specialists in the engineering discipline, which was the fourth criterion adopted in this study (See Section 5.2.4 for detailed procedures).

3.4.2 Identification of Multiword Unit Technical Vocabulary Items in the EMDC

The study employs corpus-based frequency analysis and semantic analysis to identify technical MWUs from the EMDC. Corpus-based frequency analysis was used to generate a provisional

MWU list, while semantic rating (meaning-related analysis) was used to decide whether multiword items in this list qualify as technical vocabulary. First, the *WordList Cluster* function in WordSmith Tools (Scott, 2012) was used to generate a list of MWUs in the EMDC. This study employed four frequency-based procedures to identify technical engineering MWUs: analysing word sequences of two to five units, adopting word type as the unit of counting, applying a frequency cut-off of 20 occurrences, and ensuring that at least one word from the single-word ETV list was included. While there is no universally agreed-upon frequency threshold for discipline-specific MWUs, most studies adopt a cut-off point ranging between 20–40 occurrences per million words. For example, some studies set the threshold at 25 occurrences per million words (Coxhead et al., 2017; Wood & Appel, 2014), while others use as few as five occurrences per million words (Ackermann & Chen, 2013; Benson & Coxhead, 2022). These variations reflect the specific objectives and contexts of each study.

For semantic analysis, two meaning-related procedures were followed: consulting concordance lines and referring to specialised engineering dictionaries. These procedures helped identify distinctively technical MWUs, reducing the multiword unit list to 856 items. For instance, *Gunn diode*, an example of a MWU identified, is defined as an electron device – a form of *diode* – described as a ‘two-electrode electron tube containing an anode and a cathode’ or a ‘two-terminal semiconductor electronic component’ (McGraw-Hill’s Dictionary of Engineering, 2003). Detailed information is provided in Chapter 5 (see Section 5.3.4).

Following the above criteria, an initial list of over 8,800 provisional MWUs was generated, which included many non-technical items, such as ‘*in the’,* ‘*for the’*, and ‘*to the’*. In addition, two semantic criteria were applied to further refine this list: first, assessing whether each MWU conveyed a complete meaning, and whether the meaning of the MWU relates to engineering-specific knowledge. For instance, ‘*steam generator*’ is obviously a useful technical term in power plant engineering, but its shorter counterpart '*generator tube'* might need to be checked using concordance analysis to see if it's an elliptical reference or a separate notion. In chemical engineering, '*membrane distillation*' is a comprehensive technical procedure, yet '*contact membrane*' by itself may be unclear without supporting context. Similar reasoning applies to the more general '*license plate*' versus the more specific '*license plate recognition*' computer vision challenge.

The second semantic criterion requires that a MWU should convey engineering-specific meaning to qualify as a technical MWU. Identifying such MWUs is often more straightforward than identification of single word technical items because they frequently contain known technical single word as part of the constituents. These MWUs typically fall into two categories: 1. Partially Technical MWUs: Where at least one constituent is a technical term (e.g., *wall temperature*, *average waiting time*, *mode of vibration*). 2. Fully Technical MWUs: Where all components are discipline-specific (e.g., *steam generator tube*, *electrical power output*, *welding speed parameters*).

By systematically applying these criteria supported by corpus evidence the study produced a refined list of pedagogically relevant MWUs. This approach ensures that the selected lexical bundles are both conceptually complete and discipline-specific, thereby enhancing their utility for teaching and learning engineering vocabulary.

Furthermore, the list contained many repeated two-, three-, and four-word MWUs. To address this issue, Wood and Appel's (2014) root structure method was used to refine the list, as applied and suggested in previous studies such as Biber and Barbieri (2007) and Hyland (2008). The repeated two-, three-, and four-word units were condensed into a single root structure. For instance, '*membrane distillation'* was identified as a core structure encompassing variations such as '*in membrane distillation'* and '*direct contact membrane distillation'*. Detailed procedures are discussed in Chapter 5.

3.4.3 Analysis of Frequency Distribution of Technical Vocabulary Across the Sections of the EMDC

This section provides an overview of the analysis of the distribution of technical vocabulary distribution across the five sections of masters' dissertations. Examining the distribution of TV across dissertation sections is important for understanding the role of technical language in structuring engineering academic writing and for identifying which sections require greater lexical support for learners. Previous research focuses on the distribution of various linguistics features in different sections, most reports focus on meta-discourse markers such as stance, hedges, negation, and boosters (e.g., Charles, 2006; Loghmani et al., 2020; Taymaz 2021; Sun & Crosthwaite; 2022; Xiao & Sun, 2020). In addition, the rationale for analysing the distribution of single-word technical vocabulary across different sections of engineering masters' dissertations were to provide a clearer understanding of how technical vocabulary is used within this academic genre. It also sheds light on how technical vocabulary shapes lexical organization in various sections of engineering dissertations. This systematic approach offers a comprehensive perspective on the use of technical language and academic writing practices in the field of engineering.

Table 3-2: EMDC Sub-corpora

Sub-corpora	Number of Tokens
Introduction	173,317
Literature Review	380,112
Methodology	367,366
Results and Discussion	344,860
Conclusion	64,312
Total	1,329,967

To this end, the target corpus, the EMDC, was divided into five sub-corpora: Introduction, Literature Review, Methodology, Results and Discussion, and Conclusion, as outlined in Table 3.2.

The AntWordProfiler programmes was used to analyse the distribution of single-word-unit technical vocabulary across different sections of the engineering masters' dissertations. The identified technical vocabulary list was used as the base-level list, while each sub-corpus was entered as the target corpus to run the analysis. Due to variations in the size of the different sub-corpora, the frequencies were normalized to 1,000,000 to account for these differences, as suggested by Szudarski (2018). The frequencies were normalized using the following formula:

Normalized Frequency per Million Words = (Observed Frequency ÷ Total Corpus Size) ×1,000,000

The normalized frequency percentage was then calculated as: Normalized Frequency % = (Normalized Frequency ÷ 1,000,000) × 100

These normalized frequency scores were used to report the results in Chapter 5 (Section 5.3.1).

3.5 Overview of Methodology in Study 3: Teachers' Perception of Pedagogic Usefulness and Learners' Knowledge of Engineering Technical Vocabulary

This section provides an overview of the methodology employed in Study 3. This study aims to investigate Saudi engineering students' receptive vocabulary knowledge of ETV items. It also aims to explore teachers' perceptions of the pedagogic usefulness of these technical vocabulary items. In this study, the terms teachers and lecturers are used interchangeably. The rationale of this study is to supplement the corpus data with information from end users. Finally, it aims to examine possible relationships between Saudi engineering students' receptive vocabulary knowledge and teachers' perceptions of the pedagogic usefulness of ETV lists. This study was informed by the results of the corpus analysis technical vocabulary list developed in Study 2. A good grasp of technical vocabulary is essential in developing subject knowledge and enhancing communication in specialised language (Lui & Lei, 2020; Knoch, 2014). "The specialised language of a discipline is likely to be very closely connected to a student's knowledge of that discipline" (Woodward-Kron, 2008, p. 246). Similarly, teacher evaluation of word usefulness is essential in developing pedagogic word lists for learners, especially in L2 or foreign contexts (Dang et al., 2022a). It is also important to emphasize that this approach directly supports one of the study's primary objectives: creating a practical resource for graduate students. Since the Engineering Technical Vocabulary (the ETV) list was derived from masters' dissertations which was a specialised academic genre. It is particularly relevant for its intended users at this advanced educational level. Henceforth, this study primarily intends to address the following three research questions:

1. To what extent do Saudi undergraduate engineering students receptively understand the technical vocabulary in their field of study?
2. Which technical engineering vocabulary items do teachers perceive as being useful for pedagogical purposes?
3. Is there a relationship between engineering teachers' perceptions of the usefulness of engineering technical vocabulary and the receptive knowledge of engineering technical vocabulary among Saudi undergraduate engineering students?

This section provides a summary of the study participants, data collection instruments, and data analysis procedures employed to address the aforementioned research questions.

3.5.1 Participants

The present study incorporates information from both teachers and learners to complement the corpus-based generated ETV list. Accordingly, the participants included both learner and teacher groups, as outlined below.

a. *Learner Participants:*

Seventy-eight final-year engineering students from Northern Borders University (NBU) in Saudi Arabia participated in this study. They represented four engineering disciplines: 12 from Electrical Engineering, 19 from Civil Engineering, 34 from Mechanical Engineering, and 13 from Industrial Engineering. The participants completed the Updated Vocabulary Levels Test (UVLT) (Webb et al., 2017) and the Yes/No tests, as detailed in the following sections.

b. Teacher Participants:

Twenty experienced engineering lecturers were recruited to assess the usefulness of the ETV list. Each participant had at least seven years of teaching experience in Saudi higher education, ensuring a deep understanding of the local educational context. The lecturers came from diverse educational and cultural backgrounds, offering a wide range of perspectives on engineering and teaching practices in Saudi Arabia, as fully described in Section 6.3.1.

3.5.2 Data Collection

This section provides an overview of the data collection process involving both student and teacher participants. It begins with a detailed description of the stages involved in collecting data from the student participants as follows.

3.5.2.1 Data Collection from Learner Participants

The data collection from learner participants was conducted in two stages. In the first stage, the UVLT adapted by Webb, et al., (2017) was administered to a group of 78 Saudi engineering undergraduate student participants. The UVLT was designed to measure Saudi learners' receptive knowledge of English vocabulary by providing an estimate of learners' vocabulary size (Schmitt et al., 2001). It also has been widely validated and used in educational research to categorize learners based on their vocabulary proficiency levels (Webb & Chang, 2012). This word-definition matching test is commonly used to measure receptive vocabulary knowledge (Anderson & Freebody, 1983; Harsch & Hartig, 2015, Meara & Buxton, 1987; Mochida & Harrington, 2006). The test presents words in clusters of six, comprising three target words and three distractors, along with three definitions (Schmitt et al., 2001). The rationale for using the UVLT in this study was to categorize the student participants according to their knowledge of vocabulary levels, which will inform the discussion of the results of the second stage, examining their receptive knowledge of the ETV list (See Section 6.3.2.2 for detailed information).

In the second stage, a Yes/No engineering vocabulary test was developed by the researcher and used for data collection to test Saudi engineering student participants' receptive knowledge of the ETV identified in Study 2. The test format included real words and pseudowords (Mochida, & Harrington 2006). The latter were used to limit participants' tendency to falsely claim knowledge of unknown words. Following recommendations by Meara and Buxton (1987), pseudowords were designed to equal 66.66% of the number of real words. The test comprised three blocks of 100 items each: 70 real words randomly selected from various frequency ranges and 30 pseudowords. Altogether, the test included 210 real technical vocabulary items and 90 pseudowords, providing a comprehensive measure of participants' receptive knowledge of the identified vocabulary.

3.5.2.2 Data Collection from the Teacher Participants

Three surveys utilizing a five-point Likert scale were developed to assess the pedagogical usefulness of ETV by engineering lecturers. The surveys were administered using Excel files sending to the raters by email, where lecturers rated the same 210 real words included in the Yes/No tests. These words were distributed across the three surveys based on Schmitt and Schmitt's (2014) high-, mid-, and low-frequency vocabulary bands. The first survey included 70 ETV items from the high-frequency band, the second survey featured 70 ETV items from the mid-frequency band, and the third survey comprised 70 ETV items from the low-frequency band and supplementary vocabulary. The scale ranged from 1 (least useful) to 5 (most useful) (see Appendix A). Participants received the ETV surveys in Excel files sending to the raters by email (see Section 6.3.2.2).

3.5.3 Data Analysis

This section provides an overview of the analytic procedures employed to analyse data collected from student and teacher participants in Study 3. It begins with a detailed description of the analytic procedures involved in analysing data from the student participants as follows.

3.5.3.1 Analysing the Learner Data

The data analysis from learner participants was conducted in three stages. In the first stage, the responses from 78 students' UVLT were marked and scored, with participants receiving a point for each correct response, with a maximum possible score of 30 per level. Descriptive statistics were used with the aid of SPSS version 27.0 to compute the mean and standard deviation for each level. The results of UVLT were used to categorize student participants into three groups according to vocabulary proficiency levels based on Webb, et al.'s (2017) threshold scores. The first group, classified as beginners, includes students who have not yet mastered the 2K level, scoring less than 26 out of 30 (87%) at this level. The second group, intermediates, comprises students who have not yet mastered the 3K level, achieving a score of 26 out of 30 or 87%. Lastly, the advanced group consists of students with higher-level of proficiency who have successfully mastered at least the 3K vocabulary level. To investigate differences among these groups, a one-way repeated measures ANOVA was performed. Where significant differences were identified, post-hoc tests were conducted to determine the specific levels at which the differences occurred.

The second stage involves the analysis of the Yes/No ETV test. When evaluating the results of a Yes/No test, it is crucial to recognize the four possible response categories, which can be divided into correct and incorrect responses. Correct responses include accurately identifying real words (hits) and correctly rejecting pseudowords. Conversely, incorrect responses encompass failing to recognize real words (misses) and mistakenly accepting pseudowords as genuine (false alarms) (Huibregtse et al., 2002). Among these, hit rates are thought to represent underlying vocabulary knowledge (Mochida & Harrington, 2006). (See Figure 6.2. for detailed information).

This study adopts the I_{SDT} formula to evaluate participants' responses on Yes/No tests, assessing their receptive knowledge of the ETV list. The I_{SDT} formula has proven particularly effective in addressing individual response biases (Beeckmans et al., 2001; Huibregtse et al., 2002; Mochida & Harrington, 2006). Additionally, results of other scoring methods such as calculating hits, hits minus false alarms (h-f), correction for guessing (*Cfg*), and Δm Meara's formula were reported and synthesized, providing a necessary foundation for the adoption of the I_{SDT} formula in this study (see Section 6.3.3).

Finally, bootstrapped statistical measures were employed to assess the students' receptive knowledge of engineering technical vocabulary across the three learner groups: beginners, intermediate, and advanced. These measures included central tendency indicators, variability metrics, and estimated ranges of true population parameters. Bootstrapping, a robust statistical procedure, simulates study replication by resampling from the population (Beasley & Rogers, 2009; LaFlair, Egbert, & Plonsky, 2015).

Bootstrapping was used as the primary analytical method here because of its demonstrated robustness in educational research, particularly when comparing groups with potentially uneven sample sizes or non-normal distributions (Plonsky, 2015), as is the case of the present study. By treating the study samples as a pseudo-population and generating numerous resampled datasets, this method provides reliable statistical inferences without requiring strict parametric assumptions.

Bootstrapping's particular strengths align well with the challenges inherent in vocabulary assessment research. First, it accommodates the natural variability in vocabulary acquisition patterns that often leads to non-normal score distributions across learner groups. Second, it provides more stable estimates of effect sizes and confidence intervals than traditional parametric tests when working with smaller subsamples, which is crucial for making valid comparisons between our three proficiency levels. Third, the resampling approach helps to account for potential outliers or skewed distributions that might otherwise distort our understanding of group differences.

Furthermore, bootstrapping offers several practical advantages. It allows us to determine with greater confidence whether observed differences in vocabulary knowledge scores between beginner, intermediate, and advanced learners reflect genuine proficiency gaps rather than sampling artifacts. This is particularly important when examining technical vocabulary acquisition, where exposure and mastery may vary significantly even within proficiency levels. By adopting this method, we ensure that our conclusions about the ETV list's effectiveness for different learner groups are grounded in robust statistical evidence. Further analytical procedures are provided in Section 6.3.3.5.

3.5.3.2 Analysing the Teacher Data

The second part of Study 3 focuses on evaluating the pedagogic usefulness of the ETV list by engineering lecturers. Descriptive statistics were used to analyse the lecturers' evaluations of 210 ETV items across Schmitt and Schmitt's (2014) high-, mid-, and low-frequency bands. Items were ranked based on the mean scores and standard deviations (SD) provided by the lecturers' rating of usefulness, enabling the researcher to identify the pedagogic usefulness or relative value of each item as perceived by experienced engineering educators.

Additionally, a bootstrapped Pearson product-moment correlation was used to examine the possible relationships between learners' vocabulary knowledge and teachers' perceptions of the pedagogic usefulness of the ETV list. Correlation values were interpreted based on Plonsky and Oswald's (2014) benchmarks for L2 research, with r values near .25 considered small, .40 medium, and .60 large (See Section 6.3.3.5).

3.6 Ethical Issues

This section discusses the ethical issues related to corpus research (collection of textual data) and qualitative studies involving individual participants. Adhering to ethical norms in this research helps in pursuing the objectives of the study and preventing errors that may arise (Cohen, Manion & Morrison, 2013). In conducting a corpus-based study, researchers need to consider that 'the most fundamental issue in corpus construction is whether or not you have the legal right to gather and distribute the data you intend to include in your corpus' (McEnery

& Hardie, 2011, p. 61). This research is guided by the ethical principles of conducting corpus research and the British Educational Research Association (BERA, 2018) guidelines that address informed consent, confidentiality, and anonymity across the research design.

3.6.1 Informed Consent

Informed consent is the formal process of obtaining the consent of the research participants by the researcher (Salmons, 2014). The main ethical consideration regarding informed consent faced in the present study was whether to obtain consent from each individual participant whose thesis was chosen for the first phase of the study or to gain access through the university management. This ethical dilemma was resolved at the beginning of the research when the researcher first obtained informed consent from the aforementioned university where the study will be conducted. The university management agreed to allow access to the masters' dissertations written by Saudi engineering students. These dissertations were used to form the target corpus, the EMDC. Since Study 3 of this thesis involved student and teacher participants, this necessitated another form of informed consent from the potential individual participants. The recruitment of the participants is based on British Educational Research Association (BERA) guidelines, which emphasize the principles of respect and the need for voluntary informed consent.

3.6.2 Confidentiality and Anonymity

Another important ethical consideration, as recommended by the BERA, is the protection of the participants from both physical and mental harm. The researcher also must ensure that the information obtained is truthful and kept confidential. Careful steps were taken in this study to protect the participants' privacy and to ensure the confidentiality of the data. For instance, the management of KAU instructed that all names and personal information should be removed from the collected samples. To ensure the confidentiality of the data and the participants, all names and personal information were removed from the corpus. Additionally, student and teacher participants in Study 3 were treated anonymously for the details see (Appendix C, D).

3.7 Summary of the Chapter

This summary outlines the methodology employed in conducting the current thesis. Based on the objectives of the thesis, the work was divided into three studies. Detailed information on the methodology used in each study is provided in the subsequent chapters, along with the corresponding results and discussion. Therefore, this chapter serves as a general synopsis of the methodological approach.

Chapter 4 : Study One: Lexical Profile and Vocabulary Load of the EMCD

4.1 Introduction

The present chapter outlines the procedures and findings of **Study 1**, which serves as a preliminary stage in this thesis. This study focuses on establishing the vocabulary profile of masters' dissertations in engineering, written in English by Saudi master's students. The purpose of this part of the investigation is to gain a clearer understanding of the types of words that constitute the lexicon of the masters' dissertation genre written by EFL learners and to determine the vocabulary load required to comprehend such texts.

Although the primary focus of the present thesis is the identification of engineering technical vocabulary, it was essential to conduct a profiling analysis of the Engineering Masters Dissertations Corpus (EMDC). The profiling analysis was a crucial initial step for the thesis, providing foundational data before the identification of technical vocabulary. The analysis serves two key purposes.

First, vocabulary profiling provides a descriptive overview of the general lexical characteristics of the EMDC and helps identify engineering-specific words that do not appear in the BNC/COCA base word lists (1–25) and their supplementary lists. This methodological approach is well established in previous research (e.g., Coxhead & Demecheleer, 2018; Lu, 2018; Lu & Dang, 2022) and ensures that technical vocabulary is identified against a reliable lexical benchmark.

Second, the analysis aims to provides the lexical demand of the EMDC by measuring the vocabulary load and coverage. Vocabulary load refers to the number of word families a reader needs to know to understand a text. The study examined the coverage of high-, mid-, and low-frequency vocabulary within the EMDC, using the definitions provided by Schmitt and Schmitt (2014) based on Nation's (2012) frequency lists. Establishing the coverage of these categories in student dissertations provides insight into the lexical demands for textual comprehension of these dissertations as academic texts available at the library. This analysis is a necessary step before identifying the technical vocabulary, which is the core focus of the thesis. It will also be useful for describing the coverage of technical vocabulary found in Schmitt and Schmitt's (2014) high-, mid-, and low-frequency vocabulary within the EMDC. It also informs the analysis of both single-word and multiword technical units, which will be addressed in greater depth in Chapter 5 (Study 2).

This chapter addresses the following research questions:

1. What is the lexical profile of the EMDC across Nation's (2012) BNC/COCA word frequency lists?

   a. What is the coverage of Schmitt and Schmitt's (2014) high-, mid-, and low-frequency vocabulary in the EMDC?

2. What is the vocabulary load of the EMDC across Nation's (2012) BNC/COCA word frequency lists?

The chapter is organized into four main sections. It begins with a description of the target corpus (EMDC). The second section introduces lexical profiling as the analytic approach for examining vocabulary load and coverage. The third section presents the findings related to lexical profiles and vocabulary loads in the EMDC. Finally, the chapter concludes with a summary of the findings, highlighting their implications, noting key limitations, and providing a rationale for the subsequent studies in this thesis.

4.2 The Target Corpus for Analysis

The EMDC was the target corpus for analysis in this study. A detailed description of the corpus compilation process is provided in Chapter 3 (see Section 3.2). Specifically, the EMDC was constructed by gathering masters' dissertations written by engineering students at KAU. The corpus comprises a total of 113 dissertations representing seven engineering disciplines:

Aeronautical Engineering, Chemical Engineering, Civil Engineering, Electrical Engineering, Industrial Engineering, Mechanical Engineering, and Nuclear Engineering. The refined EMDC contains 1,323,601 tokens, providing a substantial dataset for exploring the linguistic features and lexical demands of technical academic writing in engineering.

The following section outlines the analytical approach used in this stage of the study, detailing the methodologies applied to investigate the corpus and identify patterns of vocabulary coverage across engineering disciplines.

4.3 Analytic Approach: Lexical Profile Analysis

This section describes the primary analytical procedures adopted in Study 1, including frequency profile analysis and assessment of vocabulary load and coverage within the EMDC. This process involved preparing the EMDC data, adapting Nation's BNC/COCA word lists, and categorizing remaining word types to create additional band lists for items labeled as "Not in the list." As discussed in Chapter 3, lexical profile analysis was employed to gain deeper insights into the overall features of engineering vocabulary in the EMDC.

The lexical profile serves several purposes in this thesis. It provides a comprehensive overview of the vocabulary load of the EMDC, quantifying lexical difficulty by examining the distribution of high-frequency (common), mid-frequency, low-frequency, and specialised/technical words. A high vocabulary load indicates that a text contains many rare or specialised words, increasing cognitive and reading demands. In short, vocabulary load measures lexical complexity and specialization within a corpus and is commonly used to assess text difficulty and coverage in English for Specific Purposes (ESP) research, which is the focus of this study (e.g., Quero, 2015, who analyzed technical textbooks to develop medical word lists using word types).

Although vocabulary load can also be interpreted from a reader-centered perspective, such as the number of base word lists required to reach 95% or 98% lexical coverage (Nation, 2006; Webb & Nation, 2017), this study focuses on the text itself rather than reader comprehension. Accordingly, the EMDC load analysis examines the distribution of words across each BNC/COCA word list and includes words not found in Nation's (2012) base or supplementary lists. The analysis also considers the distribution of technical vocabulary not found in BNC/COCA base lists 1–25, as further explored in Chapter 5 (Study 2). It ensures that the identification of technical vocabulary in Study 2 is based on a systematically profiled corpus, where the boundaries between high-frequency, mid-frequency, low-frequency, and specialised lexical items are clearly defined.

The rationale for this sequencing is supported by previous studies that have used lexical profiling as a necessary precursor to develop specialised vocabulary lists. For example, Lu (2018) employed lexical coverage analysis in constructing a science-specific vocabulary list, while Benson (2020) and Benson and Coxhead (2022) profiled corpora of technical and applied texts to identify and validate specialised vocabulary items. These studies demonstrate that without an initial analysis of vocabulary load and coverage, attempts to identify technical vocabulary risk being incomplete or lacking in contextual precision.

Accordingly, the EMDC was profiled to analyze the coverage of its vocabulary types across the BNC/COCA framework, and to identify lexical items not included in Nation's (2012) base or supplementary lists. Additionally, the lexical profile was used to analyze the distribution of single-word engineering technical vocabulary (ETV) in the EMDC across Schmitt and Schmitt's (2014) frequency bands (high-, mid-, and low-frequency). Finally, the lexical profile analysis informed the selection of lexical items for evaluating teacher perceptions of pedagogical usefulness and assessing learners' receptive knowledge of engineering technical vocabulary in Chapter 6 (Study 3).

4.3.1 Instruments for Analysis

The target corpus of this study the EMDC was analysed using Anthony's (2014) AntWordProfiler tool, which is widely recommended software for lexical profile analysis (Đurović & et. al, 2021). AntWordProfiler enables users to assess the lexical coverage of specific base word lists within a single text or across multiple texts. Moreover, the tool provides detailed lexical coverage and frequency information for each word included in the base word lists. In the present study, the target corpus was prepared as a plain text file and uploaded into the program as a user file. After undergoing cleaning (as outlined in Chapter 3), the EMDC was uploaded into AntWordProfiler to conduct the lexical profile analysis. Nation's (2012) BNC/COCA base word lists were utilized to identify the vocabulary present in the EMDC, encompassing word families from 1,000 to 25,000 base word list along with four supplementary lists (proper nouns, marginal words, transparent compounds, and abbreviations).

Table 4-1: Nation's (2012) BNC/COCA Base Word Lists

Base word	Name	Examples
1k - 25k	Word families based on frequency and range	Play, talk, walk
31	Proper nouns	London, Smith,
32	Marginal words	Hh, phew,
33	Transparent compounds	Whitehouse,
34	Abbreviations	VIP, ER, UK, LASER

In this study, the vocabulary load of the EMDC was analysed by determining the coverage of Schmitt and Schmitt's (2014) 'high-, mid-, and low-frequency' vocabulary within the corpus. This process involved calculating the coverage of each headword in the EMDC, with headwords categorized based on their frequency in Nation's (2012) BNC/COCA frequency lists as presented in Table 4.1.

Several studies on vocabulary word lists have adopted Nation's (2012) BNC/COCA base word lists for vocabulary profiling (e.g., Lu, Boers, & Coxhead, 2021; Benson & Coxhead, 2022; Dang & Webb, 2014; Webb & Paribakht, 2015; Sun & Dang, 2020). These lists were systematically developed, spanning from the most frequent 1,000-word families to the least frequent 25,000-word families, based on their frequencies in general English. This structured framework allows for a comprehensive analysis of the vocabulary used in engineering masters' dissertations across various frequency levels.

4.3.2 Processing the Engineering Masters' Dissertations Corpus (EMDC)

After cleaning the target corpus, adapting base word lists was a crucial step in the lexical profile analysis (Nation et al., 2016). For this study, the Nation's (2012) BNC/COCA base word lists was selected. The process involved uploading Nation's (2012) BNC/COCA frequency word lists (levels 1 to 25) and supplementary word lists into the AntWordProfiler tool. The tool was originally configured with three base word lists: West's (1953) GSL (1st 1,000 and GSL 2nd 2,000) and Coxhead's (2000) AWL (570-word families). As previously stated, the EMDC was uploaded into the program as a user file after being saved in plain text format. The BNC/COCA base word lists (levels 1 to 25) and supplementary lists were used as the base word lists for the analysis.

In addition, the process involves refining the main BNC/COCA word lists by thoroughly re-evaluating words absent from the base or supplementary word lists, commonly referred to as 'Not in the list' in the literature (Benson & Coxhead, 2022; Sun & Dang, 2020). Nation, Coxhead, Chung, and Quero (2016) emphasize that revising these word lists is an ongoing effort, requiring consistent updates to maintain accuracy and relevance. The EMDC contained words not captured in the existing lists, necessitating adjustments. Nation et al. (2016) acknowledged this issue, emphasizing that low-frequency family members need to be consistently incorporated into existing word families, as the BNC/COCA word lists are still incomplete.

Preliminary results of profile analysis found 5,622-word types not covered by any existing adopted word lists, which were therefore categorized as 'Not in the list'. These words were further refined, leading to additional steps of data processing, including modifying the main BNC/COCA frequency lists and categorizing the remaining words. The words classified as 'Not in the list' within the EMDC were carefully reviewed. Any words found to belong to the existing word level lists and meet Bauer and Nation's (1993) word family criteria were added to the corresponding BNC/COCA level lists. For instance, words such as *noticeable* and *addressable* were included in base word level one, as they satisfied the derivational affixation criteria (notice + able, address + able). Further details can be found in the following section.

Table 4-2: Bauer and Nation's Seven Levels of Word Family

Level	Affixes	Example
Level 1	Every form is a different word.	*imagine*
Level 2	Regular inflections (*plural, third person singular, present tense, past tense, past particle, -ing, comparative, superlative,* and *possessive*)	**imagining, imagined, *imagines***
Level 3	Most frequent regular derivational affixes (*-able, -er, -ish, -less, -ly, -ness, -th, -y, -non*, and *–un* with restricted uses)	**imaginable,** *imaginably*
Level 4	Frequent and orthographically regular affixes (*-al, -ation, -ess, -ism, -ist, -ity, -ment, -ous*, and *in-* with restricted uses)	**imagination,** *imaginations*
Level 5	Regular but infrequent affixes (*-age, -al, -an, -ance, -ant, - ary, -atory, -dom, -eer, -en, -ence, -ent, -ery, -ese, -esque, - ette, -hood, -i, -ian, -ite, -let, -ling, -ly, -most, -ory, -ship, - ward, -ways, -wise, -ante, anti-, arch-, bi-, circum-, -counter, -en, -ex, fore-, hyper-, inter-, mid-, mis-, neo-, post-, pro-, semi-, sub-, un-*)	*imaginary*
Level 6	Frequent but irregular affixes *(-able, -ee, -ic, -ify, -ion, -ist, -ition, -ive, -th, -y, -pre, re-*)	**imaginative, imaginatively**
Level 7	Classic roots and affixes (e.g., *ab-, ad-, com-, de-, dis-, ex-, sub-)*	Not applicable

Adopted from Bauer and Nation (1993, pp. 257–262)

Table 4.2 outlines Bauer and Nation's (1993) seven-level framework for word families, illustrating the affixes corresponding to each level using the example of the word family '*imagine*' and its derived forms. It is crucial to note that when a word family is assigned to a specific level, it encompasses the base word, its inflections, and its derivations, incorporating all affixes up to that level. In developing the BNC/COCA base word lists, Nation (2012) incorporates derivational forms primarily adapted from Bauer and Nation's (1993) Levels 3 and 6. Consequently, a word family may consist of the base word (Level 1), its inflected forms (Level 2), and its derivations constructed from affixes spanning Levels 3 to 6. The subsequent sub-sections explain how the base word lists, and supplementary word lists were adapted and modified for use in the current study.

4.3.2.1 Adapting BNC/COCA Base Word Lists Levels 1 to 25

In this study, Nation's (2012) BNC/COCA frequency lists (levels 1 to 25) were refined and adapted before conducting the main profile analysis as base word lists. To achieve this, the researcher evaluated and analysed the initial set of 5,622-word types identified as 'Not in the list'. Words that aligned with Bauer and Nation's (1993) framework of word families were integrated into their respective word families and assigned to the appropriate base word levels, ranging from 1 to 25. For instance, words such as *noticeable, addressable, choosable, nonheated, reachability* and *resized* were added to base word level 1 under the headwords *address, choose, heated, notice, reach*, and *sized* because they were related in meaning and attached with the affixes appeared at Levels 3 and 6 of Bauer and Nation's (1993) scale of constructing word families. Additionally, other words that met Bauer and Nation's (1993) criteria were incorporated into their corresponding base word levels, provided their meanings were semantically related to the potential headword families and satisfied the framework established by Bauer and Nation. For example, words such as *deliverable* (deliver + able), *premixed* (pre + mixed) *nonphysical* (non + physical), *nonadaptation* (non + adaptation), *nonsmooth* (non + smooth) were added at Level 2. Words such as nondestructive (non + destructive), *nonideal* (non + ideal), noninteractive (non + interactive), nonnegative (non + negative), *resourcefully* (resource + fully), *resubmitting* (re + submitting), *retarget* (re + target), *retransmits* (re + transmits), *transmitting* (transmit + ing), recirculation (re + circulation), and *dimensionality* (dimension + ality) were incorporated into Level 3. Words such as *incompressible* (in + compressible), *conformability* (conform + ability), *parameterized* (parameter + ized), hyperparameters (hyper + parameters), *incompressible* (in + compress + able) and *independency* (in + dependency) were placed in Level 4. At Level 7 the following words were added: *condensable* (condense + able), *customizable* (customize + able) and *customization* (customize + tion). It is worth noting that some levels did not record any additions from the 'Not in the list' category.

Moreover, to ensure consistency with the BNC/COCA frequency-level lists during the processing of words not found in the lists, all hyphenated words (e.g., micro-filtration, re-processing) were thoroughly examined. Hyphens, which are often used to indicate close grammatical or semantic relationships (Nation et al., 2016) were removed during this process. The elements preceding or following hyphens can function as either affixes or independent words that are already included in the base word lists (Nation et al., 2016). In certain instances, the combined form of the hyphenated terms was identified within the BNC/COCA base word lists. For instance, *high-speed, well-known*, and *cross-section* appear as highspeed in base word level 10, *wellknown* in base word level 5, and *crosssection* in base word level 7. Similarly, when a hyphenated term involved a prefix and a base word (e.g., *pre-select*), its unhyphenated form appeared in the BNC/COCA frequency lists (e.g., *preselect*) in base word level 4. This thorough refinement process ensured the accuracy and consistency of the BNC/COCA frequency lists, establishing a dependable foundation for analysing the vocabulary in the EMDC.

4.3.2.2 Adapting the Supplementary Proper Noun List

The supplementary word base level 31 comprises a comprehensive collection of proper nouns in various forms, such as "names of people, places, countries, days, months, institutions, commercial products, and holidays" (Kennedy, 2003, p.147). While processing the 5,622-word types not present in the existing frequency lists, any identified proper nouns were systematically assigned to level 31 to ensure precise categorization during lexical profiling. These additions encompassed personal names, such as *AbdulAziz, AbdulAllah, Khaled*, and *Yusof*, and place names, including *Rabigh, Dammam, Riyadh*, and *Alyammah* University. This careful classification significantly improved the accuracy of the vocabulary analysis.

4.3.2.3 Adapting the Transparent Compounds

Base word level 33 includes transparent compound terms, such as long-term, and other multiword units. Following a review of the 5,622-word types categorized as 'Not in the list', various compound terms were incorporated into this level. These additions included terms like *pseudocode, electromechanical, headquarter, macrographs, macrostructure, multiplexer*, *multiscale*, and *multiwall*, among others. This refinement ensured comprehensive coverage of compound terms in the lexical profiling process.

4.3.2.4 Adapting the Supplementary Abbreviation List

In addition to the abbreviations already included in base word level 34, the EMDC contained numerous engineering-specific abbreviations listed under 'Not in the list'. While some appeared to be marginal letters, they were, in fact, acronyms formed by taking the initial letters of multiword expressions (Plan, 2003). Recognizing their significance and relevance within the engineering field, these acronyms were incorporated into base word level 34 to ensure comprehensive profiling. After reviewing the 5,622-word types categorized as 'Not in the list', abbreviations added to base word level 34 included NAA (*Neutron Activation Analysis*), FSS (*Frequency Selective Surface*), SKU (*Stock Keeping Unit*), NOMA (*Non-Orthogonal Multiple Access*), and WLAN (*Wireless Local Area Networks*).

4.3.2.5 Categorizing Marginal Words

Base word level 32 includes marginal words such as exclamations and letters of the alphabet (Lu 2018, Nation, 2012). This level comprised approximately 0.96% of the EMDC, a notably higher proportion than the coverage reported by Nation (2012) in the Wellington Written Corpus. A closer examination of the high-frequency lexical items within base word level 32 revealed the need for reclassification of certain terms. For example, *GHz*, representing gigahertz, a unit of frequency measuring cycles per second, was reassigned to the abbreviation list. Similarly, other items such as *map*, *kv*, *kn*, *kvp*, and *var* were also relocated to the abbreviation category for better categorization.

4.3.2.6 Creating an Engineering Base Word List

Some words that were not found in Nation's (2012) BNC/COCA frequency lists across levels 1 to 25 and in the supplementary lists of proper nouns, compounds, abbreviations, and marginal words, were inherently specialised terms used exclusively within specific fields or related disciplines (Benson & Coxhead, 2022). These words were grouped into a new category to form a specialised base word list, referred to in other profiling studies on specialised vocabulary as an additional base word list, labelled as base word 35. For example, Lu (2018) referred to such a list in the medical field as the Traditional Chinese Medicine Base Word List, while Benson and Coxhead (2022) labelled a similar list in their Spoken Rugby Corpus study as the additional base word list. Benson (2020) had earlier designated it the Rugby Base Word List.

Following a careful review of the remaining words categorized as 'Not in the list', the researcher cross-checked each potential word within the corpus and verified its meaning using dictionaries, such as the *Merriam-Webster Dictionary*. As a result, a new specialised base word list named *the Engineering Base Word List* (Base Word 35), was created. This list comprises 1,023 words out of 5,622-word types categorized as 'Not in the list'. It includes terms such as '*vorticity*' (any twisting motion in the troposphere) and '*polarizations*' (the characteristic of certain electromagnetic radiations where the direction and magnitude of the vibrating electric field exhibit a distinct and defined pattern). To ensure consistency with the existing base word lists, these words were processed as follows: the meaning of each word was verified using an engineering dictionary, and concordance lines were examined to confirm its technical usage. However, the main list of technical vocabulary used in this study was created in Study 2. The primary focus of this section was to examine the lexical load and vocabulary coverage in the EMDC.

4.4 Results and Discussion

This section presents the results of the lexical profile, vocabulary load, and coverage of Schmitt and Schmitt's (2014) 'high-, mid-, and low-frequency bands' within the EMDC, addressing the two research questions in Study 1. It also discusses the main findings that emerged from the analysis. Lexical profiling plays a crucial role in understanding the vocabulary used within a corpus, offering insights into the types of words employed. Vocabulary coverage analysis demonstrates that well-constructed word lists provide a reliable means of estimating the vocabulary level of texts and revealing their lexical characteristics. By examining what proportion of the vocabulary falls within high-, mid-, and low-frequency levels or lies outside these lists, researchers can gain valuable insights into the lexical demands of the target corpus, which indicate the writers' use of vocabulary. Such analysis also establishes a foundation for subsequent investigations, such as the development of a pedagogical word list in Study 2. This approach further highlights that vocabulary is a macro-construct that can be meaningfully divided into distinct micro-categories: high-frequency, mid-frequency, and low-frequency vocabulary, or general, academic, and technical vocabulary. Among these, technical vocabulary is of primary interest in the present research, given its centrality to the engineering disciplinary context in which this study is situated.

4.4.1 Lexical Profile of the Engineering Masters' Dissertations Corpus (EMDC)

To gain an initial understanding of the lexical characteristics of the EMDC, a lexical profiling analysis was first conducted. This step was essential for examining the overall nature and distribution of vocabulary in the corpus before moving on to the identification of technical vocabulary in later stages of the study. Given that masters' dissertations in engineering represent a highly specialised academic genre, the analysis aimed to capture the distinctive vocabulary demands of this type of writing, which combines general academic language with discipline-specific technical terminology.

The EMDC was processed using AntWordProfiler with an adapted and modified version of Nation's (2012) BNC/COCA base word lists, together with supplementary lists and the Engineering Base Word List (EBWL). This combination allowed for a more precise categorization of tokens across high-frequency, academic, and technical vocabulary bands, while also highlighting items not covered in existing lists. The outcomes of this lexical profiling analysis, summarized in Table 4.3, provide a comprehensive overview of the vocabulary composition of the EMDC and establish the foundation for subsequent analyses of engineering technical vocabulary.

Table 4-3: Lexical Profile of the EMDC across Nation's (2012) BNC/COCA Word Lists

Base word	Running words	Coverage %	Cumulative Coverage %	Type	Examples
1	783,622	59.26	59.26	3,491	*The, of, and, to, in, is, a, Figure, system*

2	196,049	14.82	74.08	3,181	*Model, results, energy, process, value*
3	148,798	11.25	85.33	3,088	*Data, temperature, analysis, network, output*
4	37,560	2.84	88.17	1,683	*Simulation, solar, parameters, grid, matrix*
5	21,374	1.62	89.79	1,148	*Membrane, voltage, thermal, generator*
6	12,428	0.94	90.73	823	*Flux, turbine, corrosion, optimization, micro*
7	10,710	0.81	91.54	637	*Antenna, multi, antennas, permeate, neural*
8	5,131	0.39	91.93	541	*Iteration, radial, hz, amplitude, ultrasound*
9	3,195	0.24	92.17	387	*Convection, irradiation, kw, entropy, speckle*
10	2,847	0.22	92.39	310	*Axial, impedance, modal, hydroelectric,*
11	2,167	0.16	92.55	282	*Modulus, mhz, broadband, diode, topology*
12	1,349	0.1	92.65	194	*Tensile, orthogonal, polynomial, torsional*
13	1,622	0.12	92.77	190	*Desalination, outage, stochastic, parametric*
14	1,335	0.1	92.87	180	*Photovoltaic, dielectric, ism, crud, jute*
15	1,200	0.09	92.96	158	*Nano, boron, pom, hysteresis, luminal,*
16	1,035	0.08	93.04	150	*Laminar, cosine, Ethernet, chiral, volumetric*
17	1,112	0.08	93.12	105	*Neutron, hydropower, neutrons, arb,*
18	373	0.03	93.15	93	*Swot, fem, polycrystalline, multipoint, login*
19	600	0.05	93.2	91	*Annulus, nanoparticles, dosimeters, isotherms*
20	320	0.02	93.22	66	*Thermos, dynamometer, neuro, cavitation*
21	151	0.01	93.23	53	*Hemodialysis, dextran, linac, chromite, ferrite*
22	263	0.02	93.25	49	*Riyals, pulsatile, baha, nuclide, digraph, hight*
23	217	0.02	93.27	45	*Penstock, deferent, bioinformatics, haar,*
24	88	0.01	93.28	34	*Coplanar, ethane, gude, keratosis, anodic,*
25	254	0.02	93.3	32	*Fl, analog, ogive, interoperability, inductance*
Proper nouns	25,024	1.89	95.19	3,991	*Saudi, Arabia, Matlab, Jeddah, Nusselt*
Marginal words	12,755	0.96	96.15	48	*E, c, mm, x, g, p, o*
Transparent compounds	6,157	0.47	96.62	593	*Bandwidth, setup, layout, biogas, feedback*
Abbreviations	37,504	2.84	99.46	3,720	*Al, pv, wimax, ghz, bls, aog, ofdm, md, db*
Engineering Base Word List	7,174	0.54	100	1,023	*Microstrip, wavelet, delamination, inverter*

Total	1,322,43 7				

Table 4.3 presents the profiling results of the EMDC across Nation's BNC/COCA frequency and supplementary base word lists (proper nouns, marginal words, transparent compounds, and abbreviations) and the additional *Engineering Base Word List.* The table shows that the first 1,000-word families accounted for 59.26% of the total running words in the EMDC, representing the largest portion of its overall coverage. As in most profiling studies (Benson& Coxhead, 2022, Lu 2018, Lu and Coxhead, 2020) the first 1,000 base-level lists contained the highest percentage of tokens in a given corpus since it comprised a list of the most frequently occurring vocabulary from the BNC/COCA. Hence, the 59.26% coverage was comparable to the 56.11% coverage reported by Lu and Coxhead (2018) in the field of Chinese Traditional Medicine. The coverage was lower compared to studies on specialised vocabulary in spoken corpora, such as Benson and Coxhead's (2022) study, where the first 1,000-word families provided 84.34% coverage in the rugby spoken corpus.

Table 4.3 also shows a sharp decrease in BNC/COCA list coverage from the first 1,000 lists to the second and the third 1,000 lists (14.82% and 11.25%). From the first 3,000 base lists, a dramatic decline in coverage was observed up to the 25th base list. This is consistent with findings from previous studies on vocabulary profiling that utilized Nation's (2012) BNC/COCA base word lists, such as those by Lu (2018), Benson (2020), and Benson and Coxhead (2022). Despite the overall decrease in coverage, there are notable exceptions among the supplementary lists, including proper nouns (1.89%), abbreviations (2.84%), and transparent compounds (0.47%).

This finding reveals that *the Engineering Base Word List* represented a substantial portion of the EMDC, accounting for 0.54% of the corpus with a total of 1,023-word types (7,174 running words). These words were specifically related to engineering, and their meanings were verified using authoritative dictionaries, such as *the Merriam-Webster Dictionary*. Examples of such words include *microstrip* (derived from Greek *mikros*, meaning 'small', and Latin *stripa*, meaning 'strip'), *rotable* (from Latin *rota*, meaning 'wheel'), and *wavelet* (from *wave* and the diminutive suffix *-let*). The list also features specialised concepts, such as *delamination* (from Latin *de*, meaning 'off', and *lamina*, meaning 'layer'), *martensite* (an eponym with the Greek suffix *-ite*), and *actuators* (from Latin *actus*, meaning 'driven'). Additionally, some terms describe physical phenomena or processes, including *anisotropic* (from Greek *an*-, meaning 'not', *isos*, meaning 'equal', and *tropos*, meaning 'turn'), *epithermal* (from Greek *epi*-, meaning 'upon', and *thermos*, meaning 'heat'), *convective* (from Latin *convehere*, meaning 'to carry together'), and *polarizations* (from Latin *polus*, meaning 'pole'). Engineering equipment terminology, such as *deaerator* (from Latin *de*-, meaning 'away from', and *aer*, meaning 'air'), also appears frequently. In previous research, Lu (2018) classified such terms as 'fully technical words' in the medical field, defining them as highly specialised lexical items used almost exclusively within specific domains. Similarly, Fraser (2013) described these words as domain-specific vocabulary, often originating from Latin or Greek. For instance, terms like '*acetylcholine*' and '*hypertension*' in Western medicine (Hsu, 2013) or pharmacology (Fraser, 2009) follow the same pattern. As in the present study, Lu (2018), Hsu (2013), and Fraser (2009) suggested these words having Latin or Greek origin can pose challenges for learners due to their length and complexity.

The profiling results displayed in Table 4.3 indicate that the cumulative coverage of the BNC/COCA base word lists (levels 1 to 25) represents only 93.3% of the tokens in the EMDC. This figure falls below the minimum threshold of 95% coverage necessary for adequate reading comprehension (Laufer, 1987). The following section discusses the lexical coverage of Schmitt and Schmitt's (2014) high-, mid-, and low-frequency vocabulary within the EMDC.

4.4.2 Coverage of High-, Mid-, and Low-Frequency Vocabulary in the EMDC

The coverage of high-, mid-, and low-frequency vocabulary in the EMDC was examined using Nation's (2012) frequency-based word lists, ranging from 1,000 to 25,000 level lists. According to Schmitt and Schmitt (2014), high-frequency vocabulary consists of the first three 1,000-word families (levels 1–3), mid-frequency vocabulary includes word families between 4,000 and 8,000 (levels 4–8), and low-frequency vocabulary encompasses those above 9,000 (levels 9–25). The overall coverage for each category was calculated by combining the coverage of the respective base word levels. This approach offers a detailed understanding of vocabulary distribution in the EMDC, as shown in Table 4.4 below.

Table 4-4: BNC/COCA Coverage of High-, Mid-, and Low-Frequency Vocabulary Across the EMDC

Frequency bands	**Base word lists**	**Coverage**	**C. Coverage**
High-frequency vocabulary (1,000–3,000)	1, 2, 3	85.33%	85.33%
Mid-frequency vocabulary (4,000–8,000)	4–8	6.6%	91.93%
Low-frequency vocabulary (9,000–25,000)	9–25	1.37%	93.3%
Proper nouns, marginal words, compounds, abbreviations	31–34	6.16%	99.46%
Engineering Base Word List	35	0.54%	100%

Table 4.4 reveals that high-frequency word families (1st, 2nd, and 3rd 1,000-word levels) accounted for 85.33% of the running words in the EMDC. Meanwhile, mid-frequency vocabulary (4,000–8,000-word levels) covered 6.6%, representing a significant portion of essential vocabulary within the EMDC. As the coverage decreased, low-frequency vocabulary (9,000–25,000) made up 1.37% of the EMDC. Furthermore, the table highlighted that, beyond Schmitt and Schmitt's (2012) high-, mid-, and low-frequency vocabulary categories, the EMDC contained a notable proportion of supplementary lists (6.16%) and engineering-specific base words (0.54%).

In sum, the lexical profile results, interpreted through Schmitt and Schmitt's (2014) framework of high-, mid-, and low-frequency vocabulary, show that high-frequency vocabulary plays a central role in the EMDC. High-frequency words account for over 85% of the corpus. This finding is consistent with earlier research, such as Lu, Boers, and Coxhead (2021) and Lu and Coxhead (2020), which reported that the first 3,000-word families contributed more than 75% of the vocabulary in a TCM English corpus, including practice texts, theoretical textbooks, and journal articles. By providing this distributional picture, lexical profiling highlights the vocabulary patterns that make up the EMDC and helps explain its overall lexical complexity as an academic genre.

Another key finding is that knowledge of proper nouns and marginal words is essential for achieving 95% lexical coverage in the engineering corpus, as discussed further in the lexical load section. The four supplementary lists, together with the specialised engineering vocabulary lists, contributed 6.16% coverage, making them the second-largest source of vocabulary in the EMDC after high-frequency words. By contrast, the additional base word list accounted for only 0.06% coverage, indicating that relatively few technical terms are entirely unique to the engineering discipline. This pattern aligns with Ha and Hyland's (2017) findings in finance, where discipline-specific vocabulary was also found to constitute only a small proportion of the overall lexicon. From a frequency perspective, this suggests that most engineering vocabulary is embedded within broader high- and mid-frequency bands, with only a limited set of items falling outside the BNC/COCA lists. A more detailed analysis of these specialised items and their characteristics is provided in Study 2.

4.4.3 Vocabulary Load of the Engineering Masters' Dissertations Corpus (EMDC) Across Nation's (2012) BNC/COCA Base Word Lists

This section addresses the vocabulary load of the EMDC across Nation's (2012) BNC/COCA frequency-level lists, supplemented by additional engineering base word lists. As highlighted in the literature, vocabulary load refers to the number of word families a learner needs to know to comprehend texts (Coxhead, 2021). Coverage refers to the percentage of word tokens in a text that learners are expected to recognize (Nation, 2013). Understanding this coverage is essential for determining the vocabulary required to achieve sufficient reading comprehension (Schmitt, Jiang et al., 2011). Vocabulary load analysis measures the lexical coverage contributed by each BNC/COCA base word list until it reaches established threshold levels of 95% or 98% coverage. These thresholds are widely acknowledged as the minimum and optimal levels of vocabulary coverage necessary for learners to adequately comprehend texts (Laufer, 1987; Laufer & Ravenhorst-Kalovski, 2010; Hu & Nation, 2000). Laufer and Ravenhorst-Kalovski (2010) argued that optimal comprehension requires 98% lexical coverage, typically achieved with knowledge of approximately 8,000-word families, while 95% coverage, corresponding to 4,000–5,000-word families including proper nouns, is considered minimal adequate comprehension. This assumes that learners can infer the remaining 2–5% of unknown words from context (Lu, 2018). Numerous studies have shown that achieving these thresholds in professional or technical texts is often challenging with general word lists alone, highlighting the importance of specialised word lists to bridge the lexical gap in specific fields.

Moreover, analyzing vocabulary load in the EMDC allows us to examine the lexical characteristics of engineering masters' dissertations as an academic genre. The EMDC reflects a combination of high-frequency general vocabulary and low-frequency, discipline-specific technical terms, showing how students integrate general academic English with specialised engineering concepts. High-frequency vocabulary contributes the bulk of tokens, indicating the reliance on common English structures for cohesion and academic expression, while technical terms and low-frequency words reveal the specialised lexical resources necessary for conveying engineering knowledge. This analysis also highlights lexical density and complexity within student writing, providing insight into the distribution of general, academic, and technical vocabulary across the corpus. By profiling the vocabulary load of the EMDC, we gain a foundation for understanding the lexical environment in which engineering students produce dissertations and for systematically identifying technical vocabulary in subsequent analyses (e.g., Lu, 2018; Benson, 2020; Benson & Coxhead, 2022).

Given the important role of supplementary lists in lexical profiling, vocabulary load is determined by combining the cumulative coverage of tokens from Nation's BNC/COCA frequency lists and supplementary lists (Nation, 2016). Consequently, the vocabulary load of the EMDC was calculated by aggregating the cumulative coverage of tokens from Nation's BNC/COCA base word lists, supplementary lists, and the newly developed Engineering Base Word List. The analysis aimed to determine the total coverage required to reach the optimal threshold of 98%, while also identifying the coverage needed to achieve the minimum threshold of 95% for reading comprehension. The EMDC was processed using the AntWordProfiler program. The results of the vocabulary load analysis are presented in Table 4.5.

Table 4-5: Cumulative Lexical Coverage of the EMDC Across the BNC/COCA Base Word Lists to Reach 95% and 98% Lexical Thresholds

Base word lists	Running words	Coverage	Cumulative Coverage
Proper nouns, marginal words, transparent compounds, abbreviations (31–34)	Supplementary lists	6.16	6.16%
Engineering Base Word List		0.54	6.7%
1	High-frequency	59.26	65.96%
2		14.82	80.78%
3		11.25	92.03%
4	Mid-frequency	2.84	94.87%
5		1.62	**96.49%**
6		0.94	97.43%
7		0.81	**98.24%**
8		0.39	98.63%
9		0.24	98.87%
10		0.22	99.09%
11		0.16	99.25%
12		0.1	99.35%
13		0.12	99.47%
14		0.1	99.57%
15		0.09	99.66%
16		0.08	99.74%
17		0.08	99.82%
18		0.03	99.85%

19		0.05	99.9%
20		0.02	99.92%
21		0.01	99.93%
22		0.02	99.95%
23		0.02	99.97%
24		0.01	99.98%
25	Low-frequency	0.02	100%

Note: The lexical coverage of 95% and 98% are highlighted in bold, at the 5th and eighth 1,000-word family levels, respectively.

Table 4.5 describes the cumulative coverage of Nation's (2012) BNC/COCA base word lists, along with supplementary lists, including proper nouns, marginal words, compounds, and abbreviations, and the Engineering Base Word List. The supplementary base word lists (levels 31 to 34) and engineering base word lists were added at the top of Table 4.5 to simplify the data presentation, following similar studies such as Coxhead, et al. (2020) and Benson (2020). The results emphasize the importance of supplementary list words in achieving 95% and 98% lexical coverage in the corpus. As Nation (2013) notes, once these words are learned, they no longer pose a significant challenge to learners (see Dang & Webb, 2014; Nation, 2006; Coxhead et al., 2017).

Results in Table 4.5 indicate that knowledge of the first 5,000-word families, in addition to the supplementary and engineering base word lists, was required to reach the minimum comprehension threshold of 95% of coverage in the EMDC. The optimal threshold of 98% coverage in the EMDC was reached at 7,000-word families plus the supplementary and engineering base word lists. This means that the comprehension of engineering masters' dissertations requires knowledge of the first five 5,000- and 7,000-word families, together with knowledge of supplementary lists and additional engineering word lists, for learners to attain minimal and optimal reading comprehension. The results indicate that engineering masters' dissertations are relatively lexically demanding. These findings align with previous studies of written discourses such as Nation's (2006) study on the vocabulary load of university textbooks, novels, and newspapers, which showed 8,000-word families plus proper nouns were required to cover 98% of the vocabulary threshold. The findings also aligned with Coxhead and Boutorwick's (2018) results, which indicated Grade 8 Math textbooks required 8,000-word families plus proper nouns, compounds, and abbreviations to reach the optimal threshold of 98% comprehension. Similarly, Coxhead et al.'s (2016) study on carpentry texts required 8,000-word families, plus carpentry-specific vocabulary not present in Nation's (2012) BNC/COCA base word lists and supplemental lists of abbreviations, proper nouns, and compounds.

In the same vein, Hsu's (2014) findings on the vocabulary load of engineering textbooks revealed that learners need around 5,000-word families to cope with the vocabulary threshold in university-level texts. Similarly, Stamatović, et al. 's (2020) study on the lexical composition of theses authored by native and non-native English speakers studying English philology reported that 95% vocabulary coverage was achieved with 4,000-word families plus supplementary lists, while 98% optimal coverage required 9,000-word families and supplementary lists. In another study concerned with highly demanding written texts, Hsu's (2013) research on medical textbooks found that a vocabulary load of roughly 14,000-word families, plus proper nouns and technical medical terminology, was required to attain optimal reading comprehension. Sun and Dang's (2020) study on English language textbooks used in Chinese high schools revealed fluctuations in the vocabulary load. They found that students need to know 3,000-word families to achieve 95% coverage and 9,000-word families to attain 98% coverage. Further emphasizing the lexical complexity of such texts, Hsu (2018) reported that for Traditional Chinese Medicine (TCM) texts, knowledge of BNC/COCA, 7,000-word families plus proper nouns were necessary to reach 95% lexical coverage, while 10,000-word families plus proper nouns were required for 98% lexical coverage.

On the other hand, the results in Table 4.5 contrast with Ward's (1999) study, which achieved 95% coverage of a corpus of foundation engineering textbooks with just 2,000-word families. Similarly, research on the lexical requirements of spoken language (e.g., Benson & Coxhead, 2022, Coxhead & Demecheleer, 2018; Webb & Rodgers, 2009) found that achieving 95% coverage required familiarity with 3,000-word families plus supplementary lists, while 98% coverage needed 5,000-word families plus supplementary lists. Benson's (2020) study, in particular, revealed that spoken rugby discourse has relatively low lexical demands. Reaching 98% coverage necessitates familiarity with the first 4,000 words, in addition to proper nouns, marginal words, compounds, and abbreviations.

If supplementary lists are excluded, the 95% and 98% vocabulary load thresholds were not achieved within the 25,000 BNC/COCA word families (see Tables 4.3 and 4.4). This finding underscores the critical role of supplementary lists in ensuring adequate text comprehension. Coxhead, et al. (2020) similarly found that 98% coverage was only achieved for the 25,000-word base list in Book 3 of a Chinese English textbook series when supplementary lists were included. This highlights the indispensable role of supplementary lists in achieving optimal lexical coverage, as 98% comprehension is rarely attainable without them. The findings further emphasize the significance of proper nouns, marginal words, and an engineering-specific word list in reaching 95% coverage of engineering masters' dissertations, contributing 6.7% of the lexical coverage, as shown in the first five rows of Table 4.5. These supplementary lists offered the second-highest coverage of the target corpus, following the high-frequency vocabulary bands (levels 1–3). Their inclusion is crucial for two primary reasons. First, their extensive coverage significantly enhances comprehension. Second, as Nation et al. (2016) observed, these words generally require minimal learning effort, as students can quickly recognize proper nouns, abbreviations, and transparent compounds, deducing their meanings with ease.

Interestingly, the engineering-specific word list contributed only 0.54% coverage, suggesting that technical words exclusive to engineering are relatively few. This finding aligns with Ha and Hyland's (2017) study on finance-specific vocabulary and Benson's (2020) research on a rugby corpus, which also found limited but highly specialised technical vocabulary. However, Chapter 5 will provide a more in-depth analysis of the broader role of engineering technical vocabulary across the 25,000 BNC/COCA word list frequency levels, as emphasized by Nation (2016).

The lexical profiling and vocabulary load analysis presented here are essential frequency approaches for evaluating whether a text is appropriate for learning (Webb & Nation, 2008). While the 98% lexical threshold was achieved at level seven, both the 95% and 98% thresholds fell within Schmitt and Schmitt's (2014) mid-frequency lexical band. This reflects the significant role of mid-frequency vocabulary in bridging the gap between high-frequency and specialised word lists, providing a foundation for effective comprehension in specialised contexts.

4.5 Summary of the Findings

This chapter examined the lexical characteristics of the EMDC, focusing on lexical profiling and vocabulary load to understand the types of vocabulary used in engineering masters' dissertations written by EFL students in Saudi Arabia. The analysis was conducted using AntWordProfiler, which allowed frequency-based profiling, including vocabulary coverage and load analysis. The results indicate that the EMDC is lexically demanding, requiring knowledge of 5,000–7,000 word families, including supplementary and engineering-specific lists, to achieve minimum (95%) and optimal (98%) lexical coverage. High-frequency words (levels 1–3) account for most tokens, forming the foundation for cohesion and academic expression, while supplementary lists, comprising proper nouns, marginal words, abbreviations, and transparent compounds, contributed 6.7% of coverage, highlighting their importance for comprehension. The engineering-specific word list contributed only 0.54%, indicating that highly specialised technical terms are relatively few but essential for conveying discipline-specific meaning.

The findings show that both 95% and 98% vocabulary thresholds fall within mid-frequency lexical bands, reflecting the pivotal role of mid-frequency vocabulary in bridging high-frequency and technical terms. From a text-centered perspective, the vocabulary load analysis aids in understanding the lexical complexity and nature of the words students use, illustrating how EFL learners educated through EMI deploy a mix of general, academic, and technical vocabulary to meet the demands of the engineering dissertation genre.

Overall, lexical profiling provides not only a quantitative measure of vocabulary coverage but also insights into the composition of words in the EMDC, establishing a foundation for the subsequent identification of technical vocabulary for engineering in Study 2. This approach was also used in similar studies on specialised corpora (Ha & Hyland, 2017; Coxhead et al., 2020; Benson, 2020), confirming the necessity of supplementary and discipline-specific lists for achieving adequate lexical coverage in specialised academic texts.

The analysis, therefore, highlights that the condensed list of MWUs in the EMDC is largely characterized by variable slots before the root structure. From a practical standpoint, this reflects the linguistic nature of academic engineering writing, where function words (e.g., articles, prepositions, auxiliaries) and other grammatical markers tend to precede technical nouns and noun phrases, shaping how meaning and precision are conveyed. Function words and prepositions are among the most frequent lexical classes in both spoken and written English, a finding repeatedly demonstrated in corpus studies (Sinclair, 1991; Biber et al., 1999; Gray & Biber, 2012). Gardner and Davies (2014), in their work on the Academic Vocabulary List, similarly showed that grammatical words form the backbone of academic discourse, occurring more consistently and with greater frequency than content words. In the EMDC, these function words frequently account for the variability seen before root structures, signaling their central role in establishing cohesion, grammatical correctness, and discourse flow.

In contrast, the variables occurring after the root structures are more often content words, such as nouns or technical terms, which expand or specify the meaning of the headword. This aligns with the nature of engineering discourse, where post-modifiers (e.g., *sources*, *tube*, *coefficient*) add precision and detail to complex noun phrases. Theoretically, this distinction between pre-root and post-root variability illustrates how MWUs in academic engineering texts balance grammatical structure with semantic specificity: the pre-root elements provide the scaffolding, while the post-root elements deliver the technical substance. This dual pattern reflects both the structural norms of English academic writing and the discipline-specific requirements of engineering discourse, making the findings valuable for understanding not only how MWUs function in specialised texts but also how EFL engineering students employ language within the constraints of EMI.

Taken together, the results highlight both practical and theoretical interest. Practically, they suggest that EFL learners in engineering should pay close attention to function words and prepositions to develop fluency in constructing technical MWUs. Theoretically, the findings contribute to our understanding of how variability in MWUs maps onto the grammatical–lexical divide, reinforcing the broader corpus-based observation that while function words dominate in frequency, content words drive disciplinary specificity. This balance is central to the construction of technical academic genres such as masters’ dissertations in engineering.

Chapter 5 : Identification Engineering Technical Vocabulary

5.1 Introduction

Building upon the lexical profiling and vocabulary load of the EMDC, this chapter focuses on the identification of single-word and multi-word units of technical vocabulary in the EMDC. As conceptualized in Section 1.6, technical vocabulary is not restricted only to words absent from level lists or low-frequency vocabulary; rather, it can encompass words listed in the high-, mid-, or low-frequency lists (Nation, 2016). Some words in GSL or AWL can assume technical meanings when they appear in specialised texts (Tongpoon-Patanasorn, 2018). This distinction is crucial, as common words can take on specialised meanings in specific contexts such as word *'flow'* in plumbing (Coxhead & Demecheleer, 2018). In this study, 'technical vocabulary' refers to both single-word technical items and multiword units (MWUs) consisting of two to five words. This study utilizes both corpus-based and semantic-based methods to identify single-word units and MWUs of engineering technical vocabulary items in the EMDC. This effective method of identifying technical vocabulary and has been adopted in most previous related studies (Benson & Coxhead, 2022; Hsu, 2018; Liu & Lei, 2020).

Using the target corpus, the EMDC, developed in the previous study, this chapter outlines the methodology used in Study 2 and presents results and discussion to address the following three research questions:

1. Which technical vocabulary items, in terms of single-word units, are used in the EMDC?
    a. How many of these belong to the high-, mid-, and low-frequency vocabulary bands, respectively?

2. Which technical vocabulary items, in terms of multiword unit items, are used in the EMDC?
3. What is the distribution of the engineering technical vocabulary list over the five different sections (Introduction, Literature Review, Methodology, Results and Discussion, Conclusion) of engineering masters' dissertations?

5.2 Identifying Single-Word-Unit Technical Vocabulary in the EMDC

During the lexical profile analysis (see Chapter 4), an exclusive technical engineering base word list (referred to as the 'not in the list') was developed. This list included words such as *wavelet* and *delamination* that were absent from the BNC/COCA (base word level 1,000–25,000) word lists. In addition to these clearly technical terms, there was a large number of technical words that appeared within BNC/COCA base word lists (base word level 1–25), including terms like *volume*, *current,* and *angle*. These words are crucial for constructing disciplinary knowledge in the field of engineering. To identify technical vocabulary within Nation's (2012) base word level 1–25, this study employed a method combining quantitative keyword analysis and qualitative analysis using a semantic rating scale. The keyword analysis helped identify potential technical vocabulary based on the assumption that technical terms occur more frequently in specialised contexts than in general English (Chung, 2003; Scott, 2006). The qualitative analysis then determined whether these candidate words were genuinely technical vocabulary.

Two important points need to be considered when creating a word list for ESP: which unit of measurement will be used to count the words and whether to concentrate on specialised terms or build the list around a common core vocabulary (Benson & Basturkmen, 2006; Coxhead, 2018). The following sub-sections outline the criteria used to identify the single-unit technical vocabulary in the EMDC. This process involved keyword analysis, frequency principles, consultation of engineering dictionaries and concordance lines, and adaptation of semantic rating scales.

5.2.1 Keywords Analysis

In this study, keyword analysis was used to generate a list of potential technical words in the EMDC. Keywords, as defined by Scott (1997), are words which occur with unusual frequency in a target corpus compared to a reference corpus. This process involves comparing the target corpus the EMDC with a reference corpus representing general English using WordSmith Tools 6.0 (Scott, 2012). This software can automatically calculate a word's keyness in the target corpus relative to the comparison corpus. The process involves three steps:

1- Use WordSmith's 'Wordlist' function to create a word list that includes every word type and its frequency for the comparison corpus.

2- Repeat the first step using the target corpus instead of the comparison corpus.

3- Compare both word lists to identify keywords in the target specialised corpus.

This study follows the procedure used by Lu (2018). Keyword analysis was chosen because it identifies potential technical vocabulary based on the assumption that technical words occur more frequently in specialised texts than in general English texts (Lu, 2018, Chung, 2003; Scott, 2006). To generate keywords in the EMDC, the BNC Written Corpus was used as a reference comparison corpus, as further explained in Section 5.2.1.1.

WordSmith Tools 6.0 (Scott, 2012) offers two statistical tests for keyword analysis: Chi-square and Log-likelihood (LLR). The LLR test demonstrates superior performance when comparing lengthy texts or an entire genre to a reference corpus (Dunning, 1993; Scott, 1998), as is the case in the present study. The process of generating a potential technical vocabulary list involves examining words whose occurrences in the technical corpus are statistically higher (positive keywords) or lower (negative keywords) relative to their occurrences in the general corpus. After uploading the target corpus, the EMDC, function words and the supplementary lists identified in Study 1 were used as stop-word lists in WordSmith Tools. Thus, a list of all possible technical vocabulary items can be generated using keywords, which are ordered according to their relative keyness. The focus here was on the keywords with positive LLR since they were statistically more probable to occur in engineering discourse. The analysis output found 2,470 keywords with a positive LLR. The ten top positive words included *figure, antenna, system, temperature, data, model, simulation, results,* and *energy.* Given the large number of keywords, especially for semantic rating purposes, a frequency principle was applied as described in Section 5.2.2 to further refine the list.

5.2.1.1 The Reference Comparison Corpus

In studies aimed at identifying specialised vocabulary, it is common practice to employ a general corpus such as the BNC/COCA as a comparison corpus (Chung, 2003; Chung & Nation, 2004; Scott, 1997). This study employs the Written British National Corpus 2014 (Written BNC, 2014) as the reference corpus for comparison with the EMDC. The Written BNC 2014 comprises around 90 million words of written British English published between 2010 and 2018, collected from diverse genres to ensure a broad representation of English across different contexts (Brezina, Hawtin, & McEnery, 2021). The written BNC2014 is composed of a wide variety of texts, including academic prose (20%), fiction (20%), newspaper texts (20%), magazines (20%), E-language (5%), and others (5%).

Since the target engineering corpus (the EMDC) is also written, this study focuses on written materials within the BNC. This approach is intended to avoid generating keywords specific to spoken language, which tend to differ significantly from written texts in terms of vocabulary

(Dang, 2017; Dang & Webb, 2014; van Zeeland & Schmitt, 2013). Using the BNC as a general written corpus is ideal for identifying technical vocabulary as it offers a baseline for comparison. It also helps pinpoint potential technical vocabulary items for further qualitative semantic analysis, ensuring that the vocabulary identified is truly specialised to the field of engineering.

5.2.2 Frequency Principles (Adopting the Frequency Cut-off Points)

The frequency principle was applied to select potential technical vocabulary for inclusion in the pedagogical technical engineering word list from the EMDC. However, it is important to note that there is no universal frequency threshold that can be used in all corpus-based word list studies (Nation, 2016). The frequency threshold depends on the goal of each investigation's word list, considering the cut-off points used in previous studies such as Coxhead's (2000) 100 occurrences in a 3.5-million-word corpus, Ha's (2015) 63 occurrences in a 6.7 million running word corpus, and Lu's (2018) 60 occurrences in a 5 million running word corpus. Therefore, in the present study, a cut-off point of 60 times occurrences was applied. This means any word which occurred with a frequency of 60 times or more in the EMDC list of positive keywords was selected for the potential engineering technical word list**.** Additionally, a cut-off point of four occurrences was applied to select the off-list vocabulary identified in the first study. While conducting the lexical profile analysis reported in Chapter 4, an additional base word list was created, including words (e.g., *microstrip, wavelet, delamination, inverter*) which did not appear in any of the BNC/COCA word lists. These words, termed 'engineering base words' in this study, are very important and clearly technical vocabulary within the engineering discipline. The threshold was reduced to four for these words due to their lower frequency. It should be noted that this principle was applied after the keyword analysis, and the resulting list underwent further assessment, as illustrated in subsequent sections. The next stage was conducted by the researcher before sending the list of potential technical words to the independent raters.

5.2.3 Consulting Engineering Dictionary and Concordance Lines

This procedure involved verifying the meaning of words by consulting multiple dictionaries, such as the McGraw-Hill Dictionary of Engineering (McGraw-Hill, 2003) and the Merriam-Webster Online Dictionary (Merriam-Webster, n.d.), in addition to confirming meanings by examining concordance lines when necessary. Of the 1,146 words identified using the frequency principle, 536 were found in the engineering dictionary. The remaining 613 words were then subjected to qualitative semantic analysis by independent raters who were specialists in the engineering discipline, as described in the next sub-section.

5.2.4 Adapting the Semantic Rating Scale for Single-word Units

After identifying potential technical engineering words based on the keyness analysis and frequency cut-off point of 60 occurrences, it was recognized that not all items in the potential words list were technical, such as *study, chapter,* and *world* (Coxhead et al., 2017). Therefore, a qualitative analysis was employed to determine whether these terms words were indeed technical. A semantic rating scale adapted from Chung and Nation (2003) was used to ensure the qualitative analysis was both valid and systematic.

The semantic scale developed by Chung and Nation (2003) has been subsequently adapted in numerous ESP studies, including those in engineering (Hsu, 2014), medical textbooks (Quero, 2015), and other fields such as plumbing (Coxhead & Demecheleer, 2018), carpentry (Coxhead et al., 2016), fabrication (Coxhead et al., 2018), and applied linguistics (Fraser, 2005). Chung and Nation's (2003, 2004) extensive study determined that employing a semantic rating scale is the most reliable approach for identifying technical vocabulary in specialised texts.

The scale was designed to differentiate between technical and non-technical words based on their semantic functions. As the study did not aim to evaluate the degree of a word's relevance to the engineering field, the semantic rating scale was modified to feature only two categories: non-technical and technical. The technical category was further divided into three parts, as shown in Table 5.1.

Table 5-1: A Modified Semantic Rating Scale Criterion

	Description	**Scale**
Non-technical	A word that is general and not related to engineering discipline. *E.g., table, show, used*	1
Technical	A word that is used in the engineering discipline but with (almost) the same meaning in general English. *E.g., power, apply, solar*	2
	A word that is used in the engineering discipline but has a different meaning in general contexts. *E.g., frequency, feed, value, breathing*	3
	A word that is unique to the engineering discipline and/or [only] associated with the engineering discourse domain. *E.g., antenna, membrane, voltage*	4

Table 5.1 indicates that words in level 1 were non-technical, while words in levels 2, 3, and 4 were technical vocabulary. The raters used scale levels 1 to 4 to rate the word list in the first file. They used these scale levels to decide if each word contained in the second file was engineering technical vocabulary or not. The updated scale differs from the Chung and Nation (2003) scale in two important ways. Part 2 of the technical scale in Table 5.2 shows the first difference. According to Step 2 of Chung and Nation's (2003) scale, this group of terms would only have a passing relationship to the specialist field. However, Chung and Nation (2003) could not provide a thorough justification for their non-technical classification. Although these terms do not accurately reflect the fundamental ideas of the engineering discipline, they were deemed technical in this study since they aid in the acquisition of subject knowledge, especially for second language learners. The second distinction is that Chung and Nation's rating scale did not contain Part 3 of the technical scale (2003).

5.2.5 Procedure for Analysing the Semantic Rating Scale

Since the researcher was not a specialist in the engineering domain, she recruited three independent raters to use the semantic rating scale to assess the potential engineering technical vocabulary items. They were all experienced lecturers in the engineering domain from different settings. One of them had ten years of working experience in a university in the United Kingdom. The second had more than ten years of working experience in a Saudi University. The last rater had more than 20 years of working experience in a Saudi University. The raters were referred to as ‘domain experts’ (Long, 2005). Domain experts were recruited as recommended by Chung and Nation (2003, 2004), who emphasized the importance of specialised knowledge in correctly classifying words. If the researcher is not well-versed in the specific field, the assistance of a subject expert is necessary.

In the same vein, other related studies used multiple raters when adopting semantic rating scales (e.g., Benson, & Coxhead, 2022; Coxhead & Demecheleer, 2018, Coxhead et al., 2016). Benson, the lead author of Benson and Coxhead’s article, had experience in Rugby. Therefore, one of the three raters adopted the semantic rating scale in the same manner without checking the meaning in the dictionary or consulting the concordance lines. On the other hand, Coxhead and Demecheleer (2018), not being experts in plumbing, enlisted the help of plumbing tutors from a polytechnic to apply a semantic rating scale to identify technical vocabulary. Coxhead et al. (2016) utilized a semantic rating scale to identify technical terms in carpentry. The scale involved three raters categorizing words into three groups: technical word e.g. *flammable, apron*, a word with minimal relatedness to carpentry e.g. *hazardous, measure*, or a general word e.g. *improve, category*. Coxhead et al. (2016) concluded that a combination of corpus comparison and semantic rating scales can be effective for small and spoken corpora, although the process is time-consuming.

Additionally, Chung and Nation (2003, 2004) cautioned that the process of classifying every word in a specialised corpus can be highly time-consuming. For example, in Quero's (2015) study on medical English, the semantic rating scale was used to identify 32,194 technical words in a medical corpus, a task that took over three months to complete. Regarding the present study, the number of potential technical vocabulary was not substantial, with 536 words out of a potential 1,146 words already identified using the engineering dictionary (as discussed in Section 5.2.3). These were excluded from the lists because they were already confirmed as technical words. Therefore, the remaining 613 were sent to the three experts for rating. These raters were asked to use the semantic rating scale to rate the keywords based on their knowledge of engineering without consulting the dictionary or concordance lines. Two files containing the potential engineering technical vocabulary list together with the semantic rating scale were sent to the raters (see Appendix A). The researcher also explained to the raters all the processes and what they were required to do for assessment.

5.2.5.1 Inter-rater Reliability

To enhance inter-rater reliability, the results should be reviewed by multiple raters (Chung & Nation, 2003, 2004). This process involves assessing the agreement between raters regarding a particular dataset. Inter-rater agreement refers to the method of estimating the reliability and consistency of coding (Brezina, 2018). In this study, Gwet's AC_1 and Fleiss's κ statistical analyses were used. Brezina (2018) identifies three types of judgement variables for calculating inter-rater agreement measures. Firstly, nominal categories with two or more values can be assessed by two raters. This can be analysed using Gwet's AC_1 and Cohen's κ statistics. If a nominal variable contains two or more values, three or more raters are required. Secondly, for the ordinal (ranks) variable, two or more values and raters are required. This can be analysed using Gwet's AC_2 statistic. Thirdly, for the interval/ratio (scale) variable, the range is used with two or more raters and analysed using interclass correlation. The researcher expected to achieve acceptable inter-rater agreements, typically ranging from 0.61–0.80 (61%–80%) for 'substantial agreement' and 0.81–1.00 (81%–100%) for 'almost perfect agreement' (Larsson et al. 2020; Landis & Koch, 1977). Scores between 0.41–0.60 (41%–60%) are considered to indicate 'moderate agreement'.

All the items that met acceptable minimal inter-rater agreement as technical vocabulary were incorporated into the list. The raters identified 394 words as engineering technical vocabulary, bringing the total number of engineering single words to 930. The overall inter-rater agreement among the three raters was 94%.

5.3 Identifying Technical Engineering Multiword Units in the EMDC

Moving forward from identifying single-word ETV, the next step was to identify multiword units (MWUs). This study adopted a mixed-method approach of corpus-based frequency analysis, followed by qualitative semantic analysis, to identify technical MWUs from the EMDC. The corpus-based frequency analysis generated a provisional multiword unit list, while the qualitative analysis determined whether multiword items in the provisional list were technical vocabulary or not. Four frequency principles were followed to identify technical engineering MWUs in the EMDC:

1. Length: Two-to-five-word units

2. Unit of counting: Word type

3. Frequency: 20 or more times in the target corpus

4. Technicality: Contains a technical word from the single engineering word list.

However, the requirement that a multiword unit must contain at least one single-word technical term may have led to the omission of certain technical or semi-technical vocabulary. Semi-technical vocabulary is hybrid in nature, conveying general meanings while also activating specialised senses that differ from everyday contexts (Le & Miller, 2025). Such vocabulary is often polysemous, which, as Fraser (2012) notes, "provides learners with the greatest difficulty" (p. 135). For instance, the phrase *on the surface* carries a specialised meaning in materials or mechanical engineering, yet it contains no isolated technical word and would be excluded by the current criteria.

Similarly, some low-frequency technical words were excluded from the single-word list due to their rarity in the EMDC. This, in turn, caused multiword units containing these words (e.g., *insulation engineering* or *truncation error*) to be omitted. Consequently, the adopted approach may underestimate the presence and importance of less frequent but highly discipline-specific vocabulary in engineering dissertations.

5.3.1 Deciding the Length of Multiword Units

Deciding on the length of the multiword items is crucial before proceeding to data-driven computational analysis. As noted in Chapter 2, this study adopted the term 'multiword units', which has been the preferred term for a specialised list (Siyanova-Chanturia & Omidian 2020, p. 530) over other similar terminologies such as lexical bundles, formulaic language (Schmitt, 2010; Wray, 2000), and formulaic expressions (Biber et al., 2004). Biber et al. (1999) suggested using three to five words in extracting multiword units. Coxhead, Dang, and Mukai (2017) used corpora of labs, tutorials, and labs plus tutorials to create three lists of 4-word word strings. The most popular method for identifying MWUs is a four-word sequence (Simpson-Vlach & Ellis, 2010; Wood & Appel, 2014). However, this study used a length of two to five words, following researchers in discipline-specific MWU studies (e.g., Benson & Coxhead; 2022;

Coxhead et. al, 2017, Lu, 2018). The inclusion of two-word units was due to some important technical MWUs being two-word unit phrases (e.g., *Gunn diode, Mac layer*). Although 2-word MWUs are usually listed in 3-, 4-, and 5-word units (Simpson-Vlach & Ellis, 2010), ways to deal with the overlapping of smaller word units into larger word units are discussed in Section 5.3.5.

5.3.2 Extracting the Engineering Multiword Units

Having determined the length and essential units for counting multiword items in the previous section, the focus shifted to extracting multiword units from the target corpus. The EMDC was uploaded into the WordSmith Tools. The *WordList Cluster* function of WordSmith Tools, which produces frequency lists of sequences of tokens, was utilized. Initially, lists of all provisional MWUs were retrieved, containing over 8,800 of MWUs. However, it was observed that the lists included some MWUs which were non-technical in nature, such as function words (e.g., *for the, to the, in the*). This indicated that the provisional multiword units were extensive and required significant further work to make them comprehensive and more manageable.

Table 5-2: The Most Frequent 20 Multiword Units

1	Of the	11	From the
2	In the	12	That the
3	To the	13	With the
4	On the	14	Is a
5	And the	15	Shown in
6	For the	16	Such as
7	Can be	17	By the
8	Is the	18	Of a
9	Number of	19	Will be
10	To be	20	At the

5.3.3 Determining Frequency Criterion for the Engineering Technical Multiword Units List

It should be noted that there is no universally agreed frequency threshold for including MWUs in discipline-specific lists. Most relevant studies adopt a frequency cut-off of 25 per million words, aligning with the common range of 20–40 per million words reported in similar research (e.g., Chen, 2010; Chen & Baker, 2010; Cortes, 2004). Some studies use cut-off points of five or more times per million words in the target corpus (e.g., Benson & Coxhead, 2022 Ackermann & Chen, 2013), while others use 20 occurrences per million words (Byrd & Coxhead, 2010; Ha, 2015; Liu, 2012) or 25 occurrences per million words (Coxhead et al., 2017; Wood & Appel, 2014), based on specific research objectives. This study employed a cut-off point of 20 occurrences per million words to reduce the provisional MWU items to a manageable list. To maintain consistency with the single-word analysis, word type was chosen as the unit of counting. The next section details the extraction procedure of MWUs from the target corpus.

5.3.4 Determining Meaning Criterion for the Engineering Technical Multiword Units List

After establishing length and frequency criteria for extracting MWUs, as detailed in the previous section, the focus shifted to two meaning-related selection criteria: completeness of meaning for each MWU and whether the meaning of the MWU is related to the subject knowledge of engineering discourse. The first meaning-related criterion was achieved by consulting the concordance lines of MWUs and checking meanings in dictionaries. For instance, *Gunn diode* is an electron device which is a form of *diode*, defined as a 'two-electrode electron tube containing an anode and a cathode' or a 'two-terminal semiconductor electronic component' (McGraw-Hill's Dictionary of Engineering, 2003). Determining the completeness of meaning for terms like *Gunn diode* required consultation of both concordance lines and dictionaries.

Left Context	Hit	Right Context
An effective heat sink for each diode is essential. Each	Gunn diode	is mounted through holes in generated microstrip line
pieces of copper. Only one DC bias to the combined	Gunn diode	is needed. The advantages of the microstrip transmission
the outside edges. In both cases the optimum position of	Gunn diode	is selected. Farther increased in the gunn of
three Gun diodes at edges. Figure 6.18.1 the Layout of three	Gunn diode	amplifiers Crystal module. ?50- ------ ------- ------- -------- --
ı antenna Crystal type Figure 6.19. the Layout (Crystal) of three	Gunn diode	amplifiers Crystal module with parasitic. 1 F Figure 6.19.2Frequ
h antenna Crystal type Figure 6.20.1the Layout (Crystal) of one	Gunn diode	amplifiers Crystal module Fig 6.20.2 Frequency Response Fig 6
ınd cascaded array of 2- terminal microstrip transmission type	Gunn diode	amplifiers with patch antennas, which is to be
with tw o Gun diodes: Fig 6.15.1 the Layout of two	Gunn diode	active antennas. Fig 6.15.2Frequency Response. Fig 6.15.3 Radi
with three (Jun diode Slots Fig 6.17.1 the Layout of three	Gunn diode	active antennas. Fig 6..2.Frequency Response Fig 6..2 Radiatior
antenna with three Gun diode Fig 6.16.1 the Layout of three	Gunn diode	active antennas. Fig 6.16.2 Frequency Response. Fig 6.16.3 Rad
which is to be investigated is using the case the	Gunn diode	can be either inserted inside the patch antenna
·oduced 25 watt output power. Therefore one can see that the	Gunn diode	can produce higher power than the 3-terminal devices

Figure 5-1: The concordance lines for the term 'Gunn diode'

The second meaning-related criterion ensured that the MWUs were related to the subject knowledge of engineering discourse. To this end, each MWU should contain a word from the single-word engineering word list as one of its constituents to be included in the provisional engineering MWU list. For instance, some technical MWUs had one technical constituent (e.g., on the *surface*, primary *coolant*, the *heat* transfer *coefficient*), while others comprised entirely technical words (e.g., *dose enhancement factor*, *epithermal neutron beam*, *antenna array(s)*). This approach helped create a more pedagogically valuable list of MWUs, focusing on both the frequency and function of the terms. The ultimate pedagogical goal of this list is to enhance specialised ESP courses, aiding learners in overcoming challenges (Benson, 2020). While it is widely agreed that corpus-based word lists are essential pedagogical resources, Martinez and Schmitt (2012) suggest considering the difficulty level of lexical items to facilitate more effective learning and teaching of technical words (Lu, 2018). Nation (2016) emphasized that a pedagogically-oriented technical word list, including the most frequent words in a specialised field, can help set learning goals and address the needs of ESP learners. Frequency-based word lists enable ESP learners to achieve communicative success with minimal vocabulary (Durrant, 2014). A frequency-based word list derived from a specialised corpus can offer critical insights into how technical words are used within that specific field (Biber, Conrad, & Reppen, 1994).

Overall, every entry which satisfied the aforementioned frequency and meaning-related criteria was selected. After this procedure was completed, a provisional list of the most distinctively technical MWUs was compiled with 856 items.

5.3.5 Refining the Engineering Technical Multiword Units List

There was considerable overlap among 856 provisional technical MWU items. In other words, many of the two-, three-, and four-word MWU items were subsumed into a single basic root structure. For example, the two-word structure *membrane distillation* was identified as a component of three- and four-word structures: *in membrane distillation*, *contact membrane distillation* and *direct contact membrane distillation*. Therefore, *membrane distillation* was categorized as the root structure. This issue has been identified in previous studies (e.g., Biber & Barbieri, 2007; Byrd & Coxhead, 2010; Hyland, 2008). However, Wood and Appel (2014) addressed this issue by introducing a procedure for identifying the root structure of each MWU and the variable slots at either of the ends of the sequence.

According to Wood and Appel (2014), whenever a shorter MWU sequence is subsumed in longer MWU sequence(s), the shorter sequence is considered the root structure, and the attached words at either end are considered variables. All variable slots were put in brackets. Here are some instances found in the EMDC: The two-word structure *negative ideal* was found to be part of other MWU sequences: (the) *negative ideal*, *negative ideal (solution), (the) negative ideal (solution*). To this end, the sequence *negative ideal* was considered as the root structure with *the* being the variable slot before the root structure and *solution* being identified as a variable after the root structure. The condensed structure was presented as *(the) negative ideal (solution*). Another example was when the *primary coolant* was embedded in five other

MWU sequences: (the) *primary coolant*, *primary coolant* (activity), *primary coolant* (specific), and *primary coolant* (specific activity). Thus, the condensed structure was presented as (the) *primary coolant* (activity/ specific/specific activity).

Table 5-3: Examples of Root Structures and the Variables

Variable	Root structure	Variable
	primary coolant	
The	primary coolant	
The	primary coolant	Activity
The	primary coolant	Specific
The	primary coolant	specific activity
Of	primary coolant	specific activity

Another criterion for refining the MWUs list was determining whether singular and plural forms had the same related meaning. Fifty-one MWUs from the 856-item list were repeated in plural forms. After consulting their meanings in concordance lines and dictionaries, all similar singular and plural types were subsumed into one root structure. A stroke and 's' (/s) were used to indicate any root structure that contained singular and plural types. For example, *desalination system* and *desalination systems, hidden layer* and *hidden layers, energy resource* and *energy resources*, *electrical load* and *electrical loads* were presented as *desalination system/s, hidden layer/s, energy resource/s*, and *electrical load/s*.

Consequently, the initial 856 MWUs were refined and condensed into a more concise list of 543 technical vocabulary items (see Appendix E). Details of the results and discussion are provided in Section 5.5.3.

5.4 Analysing of Technical Vocabulary Distribution in Different Sections of the Masters' Dissertations

One aim of the present study is to look more closely at the distribution of technical vocabulary across different sections of masters' dissertations. Specifically, this part of the study examines the coverage of the ETV list over five different sections (Introduction, Literature Review, Methodology, Results and Discussion, and Conclusion) of engineering masters' dissertations. For this purpose, the EMDC corpus was divided into five sub-corpora (see Table 3.2): Introduction Section (173,317 words), Literature Review (380,112), Methodology (367,366), Results and Discussion (344,860), and Conclusion (64,312).

The AntWordProfiler program was used with Nation's (2012) BNC/COCA 1-25,000 plus supplementary lists to determine the distribution of single-word unit technical vocabulary in different parts of engineering masters' dissertations. The list of identified technical vocabulary served as the base-level list, while each sub-corpus was entered as the target corpus.

Due to the differences in sub-corpora sizes, a direct comparison of results was not feasible. To address this issue, normalized frequency was used for reporting the findings. Normalizing frequency is essential when comparing word frequencies across datasets of different sizes (Szudarski, 2018). A common approach is to convert each frequency into a value per million or per thousand words (Szudarski, 2018). The formula for normalizing the frequency scores is: Normalized Frequency per million words = Observed frequency divided by total corpus size times 1,000,000. The normalized frequency % = normalized frequency divided by 1,000,000 times 100. This formula was used in reporting the results in Chapter 3 (Section 3.4.3).

5.5 Results and Discussion

This section presents results related to the development and distributional patterns of engineering technical vocabulary, encompassing both single-word and multi-word units. It addresses three research questions exploring different aspects of technical vocabulary in the EMDC. The analysis begins with Research Question 3, which investigates the development and distributional profile of single-word engineering technical vocabulary within the EMDC. This includes examining the alignment of these vocabulary items with Nation's (2012)

BNC/COCA base word lists to highlight their distinctiveness and relevance in technical contexts.

Research question 4 focuses on the development of engineering technical vocabulary MWUs within the EMDC. This part of the analysis investigates their development and distribution, shedding light on their functional and structural roles within the corpus. Research Question 5 explores the coverage of ETV single-word lists across the four main sections of engineering masters' dissertations: Introduction, Literature Review, Methodology, and Results and Discussion. This analysis provides insights into how technical vocabulary usage varies across different academic writing sections.

The section begins by addressing research question 3, with Sub-section 5.3.1 examining the coverage of single-word ETV in the EMDC, as benchmarked against Nation's (2012) BNC/COCA base word lists. This foundational analysis sets the stage for understanding the specialised nature of the vocabulary within the corpus.

5.5.1 Coverage of the Single-Word-Unit Technical Vocabulary in the EMDC

The first finding was related to the development of a technical engineering single-word list. These words were identified using the criteria discussed above. Initially, a Keyness analysis was conducted to identify potential technical words. The target corpus, the EMDC, was uploaded into WordSmith Tools, while function words and marginal supplementary lists were used as the stop wordlist. The initial result showed 2,472 positive keywords in the EMDC. Of these, 1,143 items satisfied the inclusion criterion of 60 occurrences as a cut-off point employed in this study (as explained in Section 5.2.4). The second frequency criterion was then adopted. This process was time-consuming, as the researcher had to verify the meaning of each word using *the McGraw-Hill Dictionary of Engineering* (2003, 2nd edition). As a result, 588 keywords were identified in the dictionary and, without meeting further criteria, directly added to the final single-word engineering word list, while the remaining 555 were not.

The 555 words not found in the engineering dictionary underwent qualitative analysis using a semantic rating scale adapted from Chung and Nation (2003). The three independent raters, specialists with over 10 years of teaching experience in engineering courses at universities, were recruited to rate these 555 positive keywords. These independent raters were regarded as 'domain experts' (Long, 2005; Benson, 2020). Prior to sending the list, the researcher consulted the raters and held separate online sessions with them regarding the rating process using Chung and Nation's (2003) semantic rating scales. A file consisting of these 555 words, arranged in alphabetical order, was sent along with the rating scale to the three independent raters. The raters were required to use their knowledge of engineering to rate each word type according to the adopted semantic rating scale (see section 5.2.4). The researcher then compared the results collected from the three raters and checked the inter-rater agreement. The list was substantially reduced to 342 technical engineering word types, which, when added to the 588-word types found in the engineering dictionary, resulted in a final single-word unit technical engineering word list of 930 words (see Appendix F). The next section presents the distribution of these 930 engineering technical words across Nation's (2012) base word lists.

Table 5-4: The Coverage of Single-Word Unit Engineering Technical Vocabulary Across Nation's (2012) BNC/COCA Base Word Lists

BNC/COCA word list	Word Type	Coverage %	Cumulative Coverage %	Examples
1	91	9.78	9.78	*Breaker, burning, bus, buses, carrier, coolant*
2	212	22.8	32.58	*Active, activity, adaptive, agents, application,*
3	267	28.71	**61.29**	*Absorption, accuracy, adoption, approximately*
4	103	11.08	72.37	*aluminium, analyser, applicable, array, arrays*
5	56	6.02	78.39	*Aggregate, altitude, amplifier, analytical, axis*
6	36	3.87	82.26	*capsule, catalyst, corrosion, fins, flux, incubators*
7	28	3.01	85.27	*Alloy, alloys, ambient, antenna, antennas, neural*
8	14	1.51	86.78	*Amplitude, attenuation, catalytic, iteration,*
9	8	0.86	87.64	*barometric, brine, convection, entropy,*
10	10	1.08	88.72	*axial, dialysis, ergonomic, hydroelectric,*
11	5	0.54	89.26	*Diode, methanol, modulus, topology, zigzag*

12	2	0.22	89.48	*Orthogonal, tensile*
13	2	0.22	89.7	*Desalination, outage*
14	2	0.22	89.92	*Dielectric, photovoltaic*
15	1	0.11	90.03	*Boron*
16	2	0.22	90.25	*Cosine, laminar*
17	3	0.32	90.57	*Hydropower, neutron, neutrons*
18	0	0	90.57	-
19	1	0.11	90.68	*Annulus*
20	0	0	90.68	-
21	0	0	90.68	-
22	0	0	90.68	-
23	1	0.11	90.79	*Penstock*
24	0	0	90.79	-
25	1	0.11	90.9	*Analog*
Proper nouns	9	0.97	91.87	*Doppler, Gaussian, Nusselt, Poiseuille, Reynolds*
Transparent compounds	10	1.08	92.9	*Aircrafts, bandwidth, ciphertext, downstream, layout*
Engineering base word list	66	7.1	100	*Actuation, actuator, alloying, anisotropic, anodization, antiscalant, antiscalants, austenite,*
	930			

Table 5.4 outlines the coverage of single-word unit ETV across Nation's (2012) BNC/COCA base word lists. Interestingly, these 930 technical vocabulary items occurred in 263,172 instances in the EMDC, representing 19.92% of the 1,320,955 total occurrences of all word types. This result indicates that single-word technical vocabulary items have substantial coverage in the EMDC. This finding aligns with Nation's (2013, p. 303) conceptualization of technical vocabulary, which he defined as words recognizably specific to a discipline that can make up a fairly large proportion of a technical text. This finding also aligns with Ward's (2009) study of engineering course books, where 299 technical words were generated, known as the Basic Engineering List, covering 16.4% of the corpus. However, this coverage is less than that found in studies on other specialised written corpora, such as Coxhead, McLaughlin, and Reid (2018), in which technical vocabulary covered 29% of the fabrication written corpus. Coxhead et al.'s (2020) study regarding trades found that technical words covered more than 30% of the total tokens in the corpus. Coxhead and Demecheleer (2018) found that technical vocabulary covered 35.58% of the plumbing written corpus. In Benson (2020), it covered 35.41% of the written rugby corpus. Chung and Nation's (2003) study found that 31.2% of the

total running words in anatomy textbooks are technical. Quero (2015) found that 37% of medical textbooks were technical vocabulary.

The relatively lower coverage of technical vocabulary in the EMDC, compared to specialised written corpora, can be attributed to key differences in the sources and purposes of the corpora. The EMDC comprises written samples produced by students, whereas previous studies primarily analysed textbooks authored by field experts. Textbooks are designed with a highly pedagogical focus, aiming to provide comprehensive and detailed content for teaching and learning. In contrast, the EMDC reflects an informative and evaluative purpose, assessing students' knowledge to fulfil academic requirements for degree qualifications in engineering disciplines. As argued by Chung and Nation (2004) and Coxhead (2018), the proportion of technical vocabulary in a corpus can significantly influence comprehension. While high technical vocabulary coverage in specialised texts is essential for expert understanding, it may present a challenge for general audiences or learners with limited background knowledge of the subject. The EMDC's moderate coverage likely reflects its dual role in both communicating technical information and meeting the evaluative needs of academic discourse in student writing.

Moreover, the 19.92% coverage of the current study is substantial compared with the coverage of technical words found in spoken corpora. For example, Coxhead, McLaughlin, and Reid (2019) found that technical vocabulary covered 9% of the fabrication spoken corpus. Coxhead and Demecheleer (2018) reported that technical vocabulary covered 11.14% of the plumbing spoken corpus. Benson and Coxhead's (2022) study on rugby found that technical vocabulary covered 12.04% of the spoken corpus.

5.5.2 Distribution of Single-Word Unit Technical Vocabulary across High-, Mid-, and Low-Frequency Vocabulary Bands

This sub-section presents the overall coverage of technical vocabulary in the EMDC across Schmitt and Schmitt's (2014) high-, medium-, and low-frequency vocabulary bands, along with Nation's (2012) supplementary lists and the Engineering Base Word List. Technical vocabulary occurs through high- (first 1,000–3,000), mid- (4,000–8,000) and low- (9,000 onwards) frequency bands in English (Nation, 2016). Table 5.5 presents the distribution

Table 5-5: Distribution of Single-Word Unit Technical Vocabulary Across Schmitt and Schmitt's (2014) High-, Mid-, and Low-Frequency Vocabulary Bands

Frequency band	Tokens	Tokens %	No. of types	Proportion of TV list	Examples
High-frequency (1,000–3,000)	207829	78.9%	570	61.29%	*Breaker, burning, bus, buses, carrier, coolant, cooling, adaptive, agents, Absorption, accuracy, adoption, architecture, atomic*
Mid-frequency (4,000–8,000)	46380	17.6%	237	25.49%	*compact, compression, configuration, amplifier, analytical, automation, Accredited, armchair, capsule, Alloy, alloys, ambient, antenna, Amplitude, attenuation, catalytic*
Low-frequency (9,000–25,000)	4569	1.73%	38	4.12%	*Antennas, Barometric, brine, convection, Axial, dialysis, ergonomic, Diode, methanol, modulus Diode, methanol, modulus, Hydropower, neutron, neutrons*
Supplementary lists	2730	1.04%	19	2.0%	*Doppler, Gaussian, Nusselt, Poiseuille, Reynolds, Aircrafts, bandwidth, ciphertext, downstream, layout, reachability,*
Engineering base word list	1915	0.73%	66	7.1%	*Actuation, actuator, alloying, anisotropic, anodization, antiscalant, antiscalants, asymptotes*
Total	263423	100%	930	100%	

Table 5.5 outlines the distribution of single-word-unit technical vocabulary items across Schmitt and Schmitt's (2014) high-, mid-, and low-frequency vocabulary bands. It also presents the technical words found outside the BNC/COCA word lists as identified in the first part of this study. This is based on the notion that technical vocabulary occurs through high-

(first 1,000–3,000), mid- (4,000–8,000) and low- (9,000 onwards) frequency bands in English (Benson, 2020; Nation, 2016; Lu 2018).

An interesting finding presented in Table 5.5 is that the majority of the technical vocabulary items (570-word types, 61.29%) were from Schmitt and Schmitt's (2014) high-frequency vocabulary band (1,000–3,000), followed by 237-word types (25.49%) from the mid-frequency vocabulary band (4,000–8,000). The low-frequency vocabulary (9,000–25,000) consisted of 38-word types (4.12%). The supplementary lists (proper nouns and transparent compounds) constituted the least frequent technical words (19-word types, 2.0%), while 66-word types (7.1%) were from the specialised engineering discipline-specific base word list. This result aligns with Benson's (2020) study, which found technical vocabulary items in both spoken (89.43%) and written (93.62%) corpora.

The high coverage of the high-frequency vocabulary based in the EMDC indicates that a considerable portion of the single-word technical vocabulary items, such as '*fault*', '*block*', and '*stress*', also have semantics functions within engineering discourse. These technical words from high-frequency band words play a crucial role in ESP, including engineering (Lu, Boers, & Coxhead, 2021). They are called '*cryptotechnical'* by Fraser (2009, p. 155). This finding was consistent with Benson's (2020) analysis of rugby corpora, which found that most technical single-word lists, both written and spoken, were from the general 3,000 high-frequency vocabulary bands (Schmitt & Schmitt, 2014). In particular, 196 (86.7%) and 223 (88.5%) of the technical written and spoken word lists, respectively, were found to be in these high-frequency ranges. As shown in disciplines including medicine (Quero & Coxhead, 2018), pharmacology (Fraser, 2005, 2009), and applied linguistics (Fraser, 2005), this lends credence to the notion that universal high-frequency language can be technical within specialised domains.

This finding emphasizes the role of general high-frequency vocabulary and illustrates how many terms are technical to the field. High-frequency vocabulary plays a crucial role in ESP, including engineering, for various reasons (Lu, Boers, & Coxhead, 2021). First, beginners in the field are likely to be familiar with many of these terms, but it is important to note that these words carry specialised meanings within the engineering discipline, differing from their general usage unknown by the learners (Coxhead & Demecheleer, 2018). Although high-frequency words may seem familiar, learners might assume they understand their contextual meaning, even though this is not always the case (e.g., Watson Todd, 2017). Due to their polysemous nature, these lexical items may have technical meanings in a specific field that are not easily understood by learners, even when they are familiar with the more general, non-technical contexts (Fraser, 2009, p. 155). As pointed out, everyday general English or academic vocabulary can be technical when it appears in specialised texts (Tongpoon-Patanasorn, 2018). Consequently, the challenge with technical vocabulary is not merely about learning new word forms and their meanings, but also, to a large extent, about establishing new form-meaning connections for words that are already familiar (Lu et. al, 2021).

Table 5.6 shows the 50 most frequent single-word unit technical vocabulary items identified from base word levels 1 to 25. Items in bold represent samples from the high-frequency vocabulary band, underlined items indicate words from the mid-frequency vocabulary band, and the remaining items are from the low-frequency vocabulary band.

Table 5-6: Top 50 Single-Word Engineering Technical Vocabulary Across Nation's (2006) BNC/COCA base word lists

SN	TYPE	FREQ.	W/BAND	SN	TYPE	FREQ.	W/BAND
1	**Figure**	7,067	High-Freq.	26	**Research**	1,273	High Freq.
2	**System**	4,729	High-Freq.	27	**Heat**	1,230	High Freq.
3	**Data**	3,260	High-Freq.	28	**Load**	1,225	High Freq.
4	**Power**	3,203	High-Freq.	29	**Frequency**	1,216	High Freq.
5	**Number**	3,166	High-Freq.	30	**Following**	1,180	High Freq.
6	**Table**	2,855	High-Freq.	31	**Low**	1,180	High Freq.
7	**Model**	2,844	High-Freq.	32	<u>Membrane</u>	1,166	Mid-Freq.

8	**Energy**	2,471	High-Freq.	33	**Type**	1,144	High Freq.
9	**Water**	2,368	High-Freq.	34	Solar	1,039	Mid-Freq.
10	**Temperature**	2,118	High-Freq.	35	**Current**	1,013	High Freq.
11	**Process**	2,011	High-Freq.	36	**Output**	1,006	High Freq.
12	**Value**	1,765	High-Freq.	37	**Parameters**	999	Mid- Freq.
13	**Performance**	1,708	High-Freq.	38	**Channel**	973	High Freq.
14	**Analysis**	1,665	High-Freq.	39	**Phase**	941	High Freq.
15	Antenna	1,541	Mid- Freq.	40	**Feed**	933	High Freq.
16	**Design**	1,509	High-Freq.	41	**Efficiency**	925	High Freq.
17	**Flow**	1,498	High-Freq.	42	Voltage	920	Mid-Freq.
18	**Average**	1,487	High-Freq.	43	**Material**	919	High Freq.
19	**Cost**	1,480	High-Freq.	44	**Input**	875	High Freq.
20	**Rate**	1,382	High-Freq.	45	**Factor**	870	High Freq.
21	**Systems**	1,361	High-Freq.	46	**Signal**	862	High Freq.
22	**Method**	1,314	High-Freq.	47	**Speed**	862	High Freq.
23	**Network**	1,314	High-Freq.	48	**Size**	854	High Freq.
24	**Effect**	1,313	High-Freq.	49	**Surface**	848	High Freq.
25	Simulation	1,302	Mid-Freq.	50	**Image**	844	High Freq.

Note: **Bold** = high-frequency vocabulary; underlined = mid-frequency vocabulary; other = low-frequency vocabulary.

As shown in Table 5.6, 44 high-frequency words such as *figure, system, data,* and *power* constituted the majority of the top 50 high-occurrence engineering vocabulary items in the EMDC. These words played a crucial role in conveying core technical meaning within the engineering discipline. In comparison, the remaining six words from the top 50 high-occurrence items, such as antenna, simulation, and membrane, were from the mid-frequency vocabulary band. No words from the low-frequency vocabulary band or other sources (e.g., supplementary vocabulary list or new-engineering TV bands) appeared among the top 50 high-occurrence engineering vocabulary items in the EMDC. Overall, these findings support Nation et al.'s (2016) assertion that engineering technical terms fall into the high-, mid-, and low-frequency vocabulary bands of the BNC/COCA word lists. Furthermore, the distribution of those technical words varied across the high-, mid-, and low-frequency vocabulary areas.

Overall, some general high-frequency words are used both in everyday contexts and within technical fields (Benson, 2020; Nation, 2012; Chung & Nation, 2003; Hsu, 2013) because every discipline shares some common ground with our daily experiences (Lu, 2018). The overlap between technical engineering vocabulary and common vocabulary is therefore understandable. For example, *bus* might only refer to a vehicle used for transportation in general English, whereas in electrical engineering it means 'a set of two or more electric conductors that serve as common connections between load circuits and each of the polarities (in direct-current systems) or phases (in alternating-current systems) of the source of electric power' (McGraw-Hill's Dictionary of Engineering, 2003). This study highlights the significant role of general high-frequency vocabulary in the specialised discipline of engineering, showing how such vocabulary is used to lexicalize technical knowledge.

5.5.3 Multiword Units Technical Vocabulary in the EMDC

Having identified the frequent single-word-unit technical vocabulary items and their coverage in the EMDC in the previous section, the focus here was on the development of MWU technical vocabulary in the EMDC. In total, a provisional list of 856 technical MWUs satisfied the inclusion criteria described in Section 5.3 (two-to-five-word units in length, word type as the unit of counting, 20 cut-off frequency point, and technicality: most contain a technical word from the single engineering word list). The initial list revealed significant overlap among two-, three-, and four-word units. To address this issue and develop a more practical MWUs list for educational use, Wood and Appel's (2014) root structure approach was implemented. This method identifies the root structure of the MWU and places any variable slots in brackets at either end of the structure (see Section 5.3.5).

The 856 MWU items were refined to a more concise list of 543 condensed MWU technical vocabulary items based on Wood and Appel's (2014) root structures, as used in other studies on ESP (Benson, & Coxhead, 2022; Benson, 2020; Coxhead et al., 2017; Lu 2018). This study supports the idea that determining the MWUs' root structure can produce a pedagogically acceptable list tailored to the needs of individual learners, reducing the learning load and improving communication (Benson, 2020). Furthermore, L2 learners can benefit from the creation of technical MWU lists, as they typically learn lexical items from lists of single words (Benson, 2020). Some technical single words that appear frequently in the word lists, such as '*imaging*' and '*Reynolds*', may lack clear meaning on their own. However, their meaning becomes more apparent in MWUs. For example, '*imaging radar'* in engineering refers to radar carried on aircraft that forms images of the terrain. '*Reynolds number'* or '*Reynolds analogy'* in chemical engineering describes the relationship showing the similarity between the transfer of mass, heat, and momentum (McGraw-Hill's Dictionary of Engineering, 2003). Reynolds number refers to a dimensionless quantity that determines the type of flow pattern of a fluid through a pipe. It was named after Osborne Reynolds who popularized its use in 1883. Building on studies that claimed technical vocabulary cannot be effectively learned in isolation (Lu, 2018; Pueyo & Val, 1996), this study reinforces the importance of integrating technical MWU lists alongside single-word technical terms in engineering language instruction. This approach will help L2 learners acquire the comprehensive knowledge needed to master technical engineering vocabulary. Table 5.7 below presents a summary of the condensed multiword unit technical vocabulary items in the EMDC is presented in Table 5.7 below.

Table 5-7: Summary of the Condensed Multiword-Unit Technical Vocabulary Items in the EMDC

Structure	Number of items	Examples
Root structures	543	(*the/polarized a microstrip/microstrip/shaped*) **Patch Antenna/s**, **the membrane** (*channel/surface*), (*the)* **heat transfer** (*coefficient/rate/in*)
Stand-alone root structures	417	<u>the natural frequencies</u>, <u>boundary condition/s</u>, <u>Reactive Power</u>, <u>composite material/s</u>, <u>energy source/s</u>, <u>energy source/s</u>
Variables before root structure	78	(*the/polarized a microstrip/microstrip/shaped*) **Patch Antenna/s,** (*out from/on*) **the system,** (*the/feed/ mass/water*) **flow rate/s,** (*the/in*) **circuit breaker/s**
Variables after root structure	44	**Renewable Energy** (*sources*), **the steam generator/s** (*tube*), **the reactor** (core), the transmission line (*effect/of*), **dose enhancement** (*factor/ factor due to*)

Variables before and after root structure	19	*(in the/the)* **power system/s** *(network), (the)* **heat transfer** *(coefficient/rate/in), (the)* **license plate/s** *(recognition), (the/green)* **supply chain (management)**, *(constant/wall)* **heat flux** *(boundary conditions/at/case), (activated)* **corrosion product/s** *(activity),*
Nominal structure	539	**the membrane** *(channel/surface),* *(the)* **heat transfer** *(coefficient/rate/in)*, **the natural frequencies, Reactive Power, composite material/s**, *(primary coolant)* **specific activity, energy source/s**
Verb structure	2	**shows the average** *(scores)*, **filtered image**
Prepositional structure	12	*(out from/on)* **the system**, **in circuit breaker/s**
Two-word structure	540	**Accuracy Measures, activated corrosion, activated CRUD, active antenna, active power**
Three-word structure	190	**wall heat flux, water flow rate, water inlet temperature**
Four-word structure	36	**in the power system**, (the) **heat transfer** (rate), (the) **heat transfer coefficient**, (the) **positive ideal solution**
Five-word structure	14	*(the)* **average number of customers**, dose enhancement (factor due to)

Table 5.7 presents a summary of the composition of condensed MWU technical vocabulary items in the EMDC with examples. The final multiword unit list consisted of 543 root structures (highlighted in bold), indicating the number of the condensed MWU technical vocabulary items in the EMDC (full list presented in Appendix E). Of these root structures, 417 were stand-alone structures underlined for easy visualization (e.g., the *natural frequencies* and *boundary condition/s*). These were root structures without any variable constituent attached at either side of the root structure.

The substantial number of technical MWUs in this study highlights the importance of MWUs and supports previous research that emphasizes their prevalence in engineering texts (Ward, 2013). Using a similar methodology, Benson (2020) identified technical MWU lists for spoken rugby, covering 239 root structures (e.g., '*over the top of*'), and 417 root structures in the written rugby corpus (e.g., '*the penalty kick is taken*'). Similarly, Lu (2018) identified 1,137 MWUs in the Traditional Chinese Medicine corpus. These findings supported the claim that technical MWUs are prevalent in specialised texts (Benson, 2020; Benson & Coxhead, 2022; Lu, 2018; Ward, 2007) and are a defining feature of such texts (Ha, 2015; Ward, 2007).

Interestingly, there were 78 variables before the root structure (e.g., (*the/polarized a microstrip/microstrip/shaped*) **Patch Antenna/s,** (*out from/on*) **the system,**), and 44 variables after the root structures (e.g., **Renewable Energy** (*sources*), **the steam generator/s** (*tube*)), with only 19 variables occurring before and after root structures (e.g., *(in the/the)* **power system/s** *(network), (the)* **heat transfer** *(coefficient/rate/in))*. All variables were italicized for ease of reading. This result aligns with prior studies on MWUs, with most variables appearing before the root structure (Benson, 2020, Byrd & Coxhead, 2010; Lu, 2018). In contrast, Benson and Coxhead (2022) found more variables after root structure than before root structure.

The analysis therefore highlights that the condensed list of MWUs in the EMDC is largely characterized by variable slots before the root structure. From a practical standpoint, this reflects the linguistic nature of academic engineering writing, where function words (e.g., articles, prepositions, auxiliaries) and other grammatical markers tend to precede technical nouns and noun phrases, shaping how meaning and precision are conveyed. Function words and prepositions are among the most frequent lexical classes in both spoken and written English, a finding repeatedly demonstrated in corpus studies (Sinclair, 1991; Biber et al., 1999; Gray & Biber, 2012). Gardner and Davies (2014), in their work on the Academic Vocabulary List, similarly showed that grammatical words form the backbone of academic discourse, occurring more consistently and with greater frequency than content words. In the EMDC, these function words frequently account for the variability seen before root structures, signaling their central role in establishing cohesion, grammatical correctness, and discourse flow.

In contrast, the variables occurring after the root structures are more often content words, such as nouns or technical terms, which expand or specify the meaning of the headword. This aligns with the nature of engineering discourse, where post-modifiers (e.g., sources, tube, coefficient) add precision and detail to complex noun phrases. Theoretically, this distinction between pre-root and post-root variability illustrates how MWUs in academic engineering texts balance grammatical structure with semantic specificity: the pre-root elements provide the scaffolding, while the post-root elements deliver the technical substance. This dual pattern reflects both the structural norms of English academic writing and the discipline-specific requirements of engineering discourse, making the findings valuable for understanding not only how MWUs function in specialised texts but also how EFL engineering students employ language within the constraints of EMI.

Taken together, the results highlight both a practical and theoretical interest. Practically, they suggest that EFL learners in engineering should pay close attention to function words and prepositions in order to develop fluency in constructing technical MWUs. Theoretically, the findings contribute to our understanding of how variability in MWUs maps onto the grammatical–lexical divide, reinforcing the broader corpus-based observation that while function words dominate in frequency, content words drive disciplinary specificity. This balance is central to the construction of technical academic genres such as masters' dissertations in engineering.

Another interesting finding presented in Table 5.7 is that the most common pattern was nominal structures (e.g., the membrane (channel/surface), (the) heat transfer (coefficient/rate/in), the natural frequencies, and Reactive Power) with a total number of 539 items. This was followed by 12 items with prepositional structures (e.g., *(out from/on)* **the system**, **in circuit breaker/s)**. The least common structure was the verb structure, with only two items occurring in the EMDC (e.g., **shows the average** *(scores)*, **filtered image)**. It should be noted that the number of variables before the root structures was considered when deciding the nominal, verb, and prepositional structures. For instance, in *(in the/the)* **power system/s** *(network),* the patterns were expanded into two: the structure with the variable *'in the'* and root structure 'power system/s' was of the prepositional pattern, while *'the* power system/s' was counted as a nominal structure. However, some root structures contain head words which could grammatically be nouns or verbs, such as *speckle, research, and transfer*. In such instances, the concordance lines were consulted to determine the forms in which such words occurred in the EMDC. For instance, the root structure '*research reactor'* includes the headword '*research*', which can function as either a noun or verb. However, the concordance lines for the phrase 'research reactor' in the EMDC indicate that 'research' is used as a noun.

ırlo Simulation platform for	research reactor	design with various r
platform was validated for	research reactor	design with two of th
ation platform OpenMC for	research reactor	design and reactor p
:s of OpenMC validation for	research reactor	design were discusse
quirements, and the current	research reactor	that already has the E
reactor facility a number of	research reactor	that has been used fc
'an [64]. There is another 56	research reactor	that is currently being
sibility of the TRiGA Mark II	research reactor	in Bangladesh to be ı
:s are running in a modified	research reactor	in Japan, the United S
ıdies that used OpenMC on	research reactor	physics calculation. 4
cellent in validating nuclear	research reactor	physics calculations. !
ong. Feasibility study of the	research reactor	being used for BNCT

Figure 5-2: Example of the concordance lines for the term 'research reactor'

Interestingly, nominal structures constituted the most common patterns of engineering MWU structures, highlighting their substantial and significant function in engineering texts. This finding aligns with Lu's (2018) study, which found that discipline-specific compound nominal phrases are essential features of Traditional Chinese Medicine (TCM) texts. These phrases often serve to name complex phenomena, entities, or processes (Ward, 2007). Similarly, Salager (1983) observed frequent use of compound noun phrases in medical English. Ward

(2007) developed a multiword item list called the ‘Technical Collocations List’, which found 78 noun phrase collocations (4 with four words, 19 three words, and 55 two words). In contrast to the present study, Ward (2007) employed only a frequency cut-off point, focusing on noun phrases that appeared three times or more in the corpus to be selected for inclusion in the list.

The prevalence of nominal structures in engineering texts can be attributed to the descriptive nature of scientific language, which relies heavily on concept-expressing nouns and nominal phrases. For example, the term *power* combines with other words to form units like *power consumption*, *power factor*, *power flow*, *power generation*, *power grid*, *power network*, *power sources*, and *power stations*. These noun combinations are key resources in constructing disciplinary knowledge (Pueyo & Val, 1996). Nominal structures are particularly effective in constructing complex engineering concepts, especially when combined with proper names, such as *Reynolds number, Nusselt number, Doppler radar,* and *Gaussian weighing method*. This finding supports the claim that technical MWUs can construct specialised knowledge in ways that individual words cannot (Lu, 2018). For example, the stand-alone words Reynolds, Nusselt, Doppler, or Gaussian lack the technical depth conveyed in their respective MWUs. This supports Pueyo and Val’s (1996) claim that some technical terms are best learned as part of broader structures rather than in isolation.

Table 5.7 shows that prepositional structures followed the noun structure, with ‘verb structure’ being the least common form in the EMDC. This contrasts with Lu’s (2018) findings, where verb structures followed the noun structures, then prepositional structures. Notably, the majority of the MWU patterns were two-word structures, occurring in 540 structures. This was followed by three-word units, consisting of 190 structures (from expanded structure). For instance, the pattern ‘the natural frequencies’ was a stand-alone three-word root structure, and *renewable energy (sources)*, *(the) received signal*, and *(the/laminar) burning velocity* were condensed three-word structures. There were 36 four-word structures, some of which were stand-alone root structures (e.g., **in the power system**), while others were condensed structures from two words with variable slots at either or both of the ends of the structure (e.g., *(the)* ***heat transfer*** *(rate)*), or consumed three-word structures with one variable slot (e.g., *(the)* ***heat transfer coefficient****, (the)* ***positive ideal solution)*****).** There were **only** 14 five-word structures, with patterns from a two-word root structure with variable slots (e.g., *dose enhancement (factor due to)*), or a three-word root structure with variable slots (e.g., *(the) average number of customers*). Benson (2020) found that the majority of root structures are two words. Of the 174 stand-alone MWUs in the word list, 154 are two-word units. As reported in previous studies, a substantial number of two-, three-, and four-word MWUs are subsumed in the three-, four-, and five-word forms. They were subsumed into condensed forms to solve the issue of overlapping (Benson, 2020; Benson, & Coxhead, 2022; Coxhead, 2019; Byrd & Coxhead, 2010; Wood and Appel, 2014).

Table 5.8 presents the top 20 most frequently refined MWU technical vocabulary items after refinement, ordered by their frequency of occurrence in the EMDC. The root structure of each multiword unit is in bold, and the variable slots are presented in brackets to enhance visibility. Additionally, the stand-alone root structures are presented in bold and italics.

Table 5-8: The 20 Most Frequent Condensed Multiword Unit Technical Vocabulary Items in the EMDC

S/No	Multiword Units	No. of Occurrences
1	(out from/on) **the system**	979
2	(the/polarized a microstrip/microstrip/shaped) **Patch Antenna/s**	503
3	**the membrane** (channel/surface)	426
4	(the) **heat transfer** (coefficient/rate/in)	370
5	***the natural frequencies***	370
6	(the/feed/ mass/water) **flow rate/s**	309
7	(in the/the) **power system/s** (network)	275
8	(the/in) **circuit breaker/s**	265
9	**Renewable Energy** (sources)	254
10	(the) **PV system/s**	244
11	(the) **radiation pattern/s**	227

12	(SF6) **circuit breaker**	223
13	**the steam generator/s** (tube)	212
14	(in/contact/ direct contact) **membrane distillation**	195
15	(the) **license plate/s** (recognition)	187
16	(the) **electrical power**	186
17	(the) **Nusselt number/s**	185
18	(the) **thrust force** (values)	180
19	**the reactor** (core)	169
20	**PDZ domain/s**	166

Table 5.8 shows that the condensed MWU technical vocabulary items formed three categories of technical MWUs in the engineering discipline: Category 1 consisted of function words as one of the constituents (e.g., (out from/on) **the system,** *the* membrane (channel/surface), (the) **radiation pattern/s**); category 2 consisted of one of the constituents being an abbreviation with technical words (e.g., (SF6) **circuit breaker, PDZ domain/s**); and category 3 had no constituent function words or abbreviations (e.g., **Renewable Energy** (sources)).

Overall, the engineering technical MWUs list further reflects the lexical complexity of engineering discourse. The full list of the technical MWUs in each category can be found in Appendix E.

5.5.4 Coverage of Single-Word Unit Technical Vocabulary Across Sections of Engineering Masters' Dissertations

This section aims to examine the coverage of the technical vocabulary in the five sections of engineering masters' dissertations (Introduction, Literature Review, Methodology, Results and Discussion, and Conclusion). Dissertations, as a form of academic genre, require writers to exhibit competence in conducting original research and the ability to produce high-quality pieces of writing to be admitted to a disciplinary academic community (Paltridge & Starfield, 2020; Sun & Crosthwaite, 2022). While a substantial amount of corpus research on dissertations and theses reported in the literature investigate various linguistics features in different sections, most reports focus on meta-discourse markers such as stance, hedges, negation, and boosters (Charles, 2006; Loghmani et al., 2020; Taymaz 2021; Sun &

Crosthwaite; 2022; Xiao & Sun, 2020). To the best of the researcher's knowledge, there is a lack of research investigating the distribution of technical words across the sections of master dissertations or PhD theses. To this end, the present study is a preliminary attempt to investigate the coverage of engineering technical words coverage across sections of engineering masters' dissertations.

Sub-corpora were created from the EMDC, each comprising words contained in the five sections of the dissertations (Introduction, Literature Review, Methodology, Results and Discussion, and Conclusion) and uploaded into the AntWordProfiler with 930 single-word ETV as base-level lists. The sub-corpora were uploaded as target files, and the analysis was run. Table 5.8 presents the distribution of the engineering technical vocabulary across the sections of masters' dissertations.

Table 5-9: Distribution of the Engineering Technical Vocabulary Across the Sections of Engineering Masters' Dissertations

Dissertation Section	No of Hits (Tokens)	Normalized Hits (per 1,000,000)	Hits (Tokens %)	Examples
Introduction	TV: 32018	184736	18.47%	*absorption, accuracy, chemical*
	Other: 141299	815263	81.53%	
Literature Review	TV: 70091	184396	18.44%	*bus, carbon, carrier, catalyst*
	Other: 310021	815604	81.56%	
Methodology	TV: 76511	208269	20.83%	*absorption, accuracy, activated, buses*
	Other: 290855	791730	79.17%	
Results and Discussion	TV: 73996	21455	21.46%	*affect, aircraft, curve, cylinder, fault*
	Other: 270864	785431	78.54%	
Conclusion	TV: 11936	185595	18.56%	*annulus, antenna, beam, boundary, catalysts,*
	Other: 52376	814404	81.44%	

Table 5.8 shows that ETV accounted for 18.47% (32,018 tokens) of the Introduction section, out of a total of 173,317 running words in this sub-corpus. Similarly, technical vocabulary

constituted 18.44% of the total running words in the Literature Review section (380,112 tokens). In the Methodology section, technical vocabulary represented 20.83% (76,511 tokens) of the total running words (367,366 tokens). The Results and Discussion section contained the highest proportion of technical vocabulary, covering 21.46% (73,996 tokens) of the total running words (344,860 tokens). Finally, the Conclusion section contained 18.56% (11,936 tokens) of the total running words (64,312 tokens).

These findings highlight the substantial role of specialised language in engineering texts, particularly in communicating research procedures and presenting findings (in Methodology and Results and Discussion sections). Conversely, the results suggest that while technical vocabulary is crucial for contextualizing one's work within the broader research landscape in the Introduction, Literature Review, and Conclusion sections, there is also a tendency to use more accessible language when framing the study's context and summarizing its implications. Finally, mastery of technical vocabulary is essential for effectively articulating core scientific content in engineering research.

Table 5-10: Distribution of Technical Vocabulary of High-, Mid-, and Low-Frequency in Each Section of Dissertations

	Dissertation Sections				
	Introduction	Literature	Methodology	Results and Discussion	Conclusion
Frequency band	Proportion of word list	Proportion of word list	Proportion of word list	Proportion of word list	Proportion of word list
High-frequency (1,000–3,000)	25,848 tokens 14.91%	54,779 tokens 14.41%	59,856 tokens 16.29%	58,829 tokens 17.06%	9,518 tokens 14.8%
Mid-frequency (4,000–8,000)	5,178 tokens 2.99%	12,348 tokens 3.25%	13,800 tokens 7.76%	13,084 tokens 3.79%	2,069 tokens 3.22%
Low-frequency (9,000–25,000)	515 tokens 0.3%	1,559 tokens 0.41%	1,284 tokens 0.35	1,055 tokens 0.31%	171 tokens 0.27%
Proper nouns and compound lists	274 tokens 0.15%	689 tokens 0.18%	987 tokens 0.27%	671 tokens 0.19%	131 tokens 0.20%

Eng. base word list	203 tokens 0.12%	716 tokens 0.19%	584 tokens 0.16%	357 tokens 0.11%	47 tokens 0.07%
Total	32,018 tokens 18.47%	70,091 tokens 18.44%	76,511 tokens 20.83%	73,996 tokens 21.46%	11,936 tokens 18.56%

Table 5.9 describes the distribution of ETV according to their levels and proportion of Schmitt and Schmitt's (2014) high-frequency vocabulary band (base word levels 1–3), mid-frequency vocabulary band (base word levels 4–8) and low-frequency vocabulary band (base word levels 9–25), plus supplementary lists and engineering base word lists in each section of masters' dissertations.

The findings show that a large proportion of ETV is from high-frequency vocabulary in the Methodology sections (59,856 tokens, 16.29%), and Results and Discussion sections (58,829 tokens, 17.06%). The prevalence of high-frequency vocabulary has two implications: (1) technical vocabulary items from the high-frequency band are important for students in reporting findings, and (2) these words are crucial for clearly articulating research procedures. Despite belonging to the high-frequency band, these words require specialised knowledge to fully grasp their contextual meaning (Nation, 2016; Lu, 2018) due to their specialised function within technical texts. As previously discussed, the EMDC contained a substantial number of high-frequency vocabulary items, which is consistent with findings from previous studies such as Benson (2020), Lu, Boers, and Coxhead (2021), Lu and Coxhead (2020), Lu (2018), and Nation (2016).

Surprisingly, Table 5.9 shows that mid-frequency vocabulary provided higher coverage in the Methodology section, representing 7.76% (13,800 tokens), followed by the Results and Discussion section at 3.79% (13,084 tokens), with the least representation in the Introduction section at 2.99% (5,178 tokens). These findings support Nation's (2013) claim that non-native English speakers should explicitly focus on learning the third to fifth 1,000-word families of mid-frequency vocabulary after mastering high-frequency words. The mid-frequency vocabulary in the EMDC appears to align with this recommendation, similar to the patterns observed by Lu and Coxhead (2017) in Chinese textbooks.

As shown in Table 5.9, low-frequency vocabulary accounted for the smallest proportion of coverage across all sections of dissertations: Introduction (0.3%, 515 tokens), Literature Review (0.41%, 1559 tokens), Methodology (0.35%, 1284 tokens), Results and Discussion (0.31%, 1055 tokens), and Conclusion (0.27%, 171 tokens). This finding also supported Lu's (2018) observation that low-frequency vocabulary contributes a relatively small yet useful amount of coverage. The low density of low-frequency words may pose challenges for learners, as inferring their meanings from context becomes difficult when these words appear only once (Nation, 2013). Similarly, the proper noun, transparent compound, and engineering-based word lists provided low coverage in all sections of the masters' dissertations. However, the Methodology section and Results and Discussion section provided higher coverage than other sections.

Overall, the distribution of technical vocabulary across different sections of dissertations remains an underexplored area. This study contributes to filling that gap by providing insights into the lexical demands of technical terms in dissertation writing. The findings highlight the need for a data-driven experimental approach to further investigate how technical vocabulary is used and the vocabulary load required in various dissertation sections. Such research could offer valuable guidance for dissertation and thesis writers in mastering the use of technical language in their academic work.

5.6 Summary of the Findings and Rationale for the Next Chapter

This chapter has provided key concepts related to procedures for identifying technical vocabulary and generating word lists. It described the different criteria for identifying single-word and MWU technical vocabulary. In summary, two pedagogical lists were developed: the first contained 930 single-word technical vocabulary items, and the second constituted 543 condensed MWU technical vocabulary items in the EMDC. The findings of this study should be interpreted with caution. While the lexical coverage of all BNC/COCA frequency levels is a key factor, semantic meaning is crucial and may also influence the comprehension of technical vocabulary (Coxhead et al., 2016). Words with multiple meanings can be particularly confusing for L2 learners unfamiliar with their technical usage. Research on technical vocabulary has consistently highlighted the challenges posed by words with multiple meanings (Coxhead et al., 2016; Lu & Coxhead, 2017; Watson Todd, 2017).

This study examined the coverage of both single-word and multiword unit technical vocabulary in the EMDC. These vocabulary lists would be pedagogically valuable for learners in the engineering field. The findings supported the idea that technical engineering word lists are essential tools for addressing L2 learners' needs. Moreover, the inclusion of high-frequency words, semantic considerations, and technical word lists advocates the importance of discipline-specific vocabulary lists in meeting the needs of ESP learners in specialised domains (Benson, 2020; Lu, 2018). These technical engineering word lists can help L2 learners focus on the technical meanings and contextual usage of these terms within the discipline (Nation, 2016). This study provided further evidence that technical word lists are valuable resources for directing learners' attention toward the most relevant vocabulary in their field. By targeting the most essential terms, discipline-specific word lists offer a practical solution to enhancing vocabulary acquisition for L2 learners in specialised domains (Durrant, 2013; Nation, 2016).

Finally, the findings of this study highlighted common linguistics features by presenting the distribution of technical vocabulary across sections of engineering masters' dissertations. Educators may take these features into account when guiding students to develop awareness of writing conventions and during thesis-writing instruction. Future research involving data-driven experiments with master and PhD students could offer deeper insights into the vocabulary used in dissertation and thesis writing.

The next chapter will present the findings of Study 3, which aims to achieve two primary objectives. First, it seeks to validate the list of engineering technical vocabulary items by examining their pedagogical usefulness through teachers' evaluations. Second, it investigates Saudi engineering students' receptive knowledge of the technical vocabulary in engineering.

Chapter 6 : Teachers' Perceptions of the Usefulness of Engineering Technical Vocabulary and Learners' receptive Knowledge

6.1 Introduction

This chapter outlines Study 3, which has two primary objectives. The first is to investigate the receptive knowledge of the engineering technical vocabulary (ETV) list among Saudi engineering students. The second is to explore engineering teachers' perceptions of the pedagogical usefulness of the ETV list developed in the previous chapter.

Vocabulary knowledge is a multifaceted and essential component of language research (Richards, 1976; Schmitt, 1998; Nation, 2013). It plays a crucial role in language teaching, learning, and achieving proficiency in L2 contexts (Trang, Truong, & Ha, 2023). Its significance in ESP has been highlighted in various studies (Dang, et al., 2022b), as discussed in Section 2.2. While different researchers may conceptualize vocabulary knowledge differently, it is widely accepted that knowing a word extends beyond recognizing its form-meaning relationship (Schmitt, 2010). Schmitt (2010) extends the argument on breadth and depth of vocabulary knowledge. While the relationship between these constructs remains a subject of debate, regression analyses suggest that depth typically provides additional explanatory power beyond size. However, the extent to which these dimensions interact depends on how they are conceptualized and measured (Schmitt, 2010).

For receptive purposes, understanding the form-meaning link may suffice. When encountering words in speech or writing, users must recognize the form and recall its meaning. Additional knowledge, such as parts of speech, derivations, and collocations, enhances comprehension but may not always be necessary to extract meaning. A depth of knowledge test measures a learner's comprehensive understanding of word attributes, including meaning, collocation, and usage appropriateness (Schmitt, 1998; Nation, 2001). In contrast, vocabulary breadth tests, like the Updated Vocabulary Levels Test, measure receptive knowledge by assessing the ability to recognize a word's form and recall its meaning during reading or listening (Nation, 2013; Schmitt, 2010).

6.1.1 Empirical Insights on Receptive Vocabulary Knowledge

Empirical studies in L2 contexts have shown that learners are more familiar with high-frequency words than lower-frequency ones (e.g., Matthews & Cheng, 2015; Nguyen & Webb, 2017). Experimental research has further revealed that word frequency influences L2 learners' language processing (Ellis, 2002; Hernández, Costa, & Arnon, 2016). For instance, Dang et al. (2022a) investigated Vietnamese EFL learners' receptive knowledge of 973 high-frequency words using a Yes/No test. Their results showed that the BNC/COCA 2000-word list provided a more reliable and pedagogically useful benchmark for L2 learners than the New-GSL, particularly because it captured a wider range of frequent vocabulary items that learners were more likely to know. This finding highlights the importance of carefully selecting the most effective general word lists for assessing receptive vocabulary knowledge, as the choice of list can significantly influence learners' measured proficiency and instructional priorities. The NGSL and BNC/COCA lists both represent high-frequency words in the corpus, but they were developed using different principles and datasets. Therefore, the decision on which list to adopt largely depends on the purpose of the study, the research design, or the specific learning objectives being pursued.

At the same time, while such studies provide valuable insights into learners' knowledge of general high-frequency vocabulary, they also underscore a significant research gap: relatively few investigations have examined learners' knowledge of technical vocabulary in L2 academic contexts. Technical vocabulary is central to mastering disciplinary discourse, particularly in fields such as engineering, medicine, and law, where comprehension relies heavily on specialised terms not typically included in high-frequency word lists. Without systematic attention to technical vocabulary, our understanding of L2 learners' vocabulary demands remains incomplete, leaving an important gap between general lexical knowledge and the specific requirements of academic and professional genres.

To address this gap, the present study investigates Saudi engineering students' receptive knowledge of technical vocabulary. It aims to provide insights into their ability to comprehend and use technical terms effectively, demonstrating their understanding of engineering content.

The second objective of this study focuses on exploring engineering lecturers' perceptions of the pedagogical usefulness of the ETV list developed in Chapter 5. Teacher evaluations of word usefulness play a key role in assessing vocabulary lists (Dang et al., 2022a). In L2 vocabulary acquisition, particularly in EFL contexts, teachers are pivotal (Dang et al., 2022a; Laufer, 2003; Schmitt, 2008). Their active participation in teaching provides them with valuable insights into the vocabulary essential for effective communication in specific contexts (Dang et al., 2022a; Laufer, 2003; Schmitt, 2008). Research has highlighted the value of teachers' perceptions in assessing vocabulary. For instance, studies in various languages, including French, have shown that teacher evaluations alone (Tidball & Treffers-Daller, 2008) or combined with lexical coverage metrics (Bardel, Gudmundson, & Lindqvist, 2012) can be more effective in gauging lexical complexity in L2 learners' speech than relying solely on corpus-based lexical coverage. Additionally, teacher feedback has been used to develop and validate academic vocabulary lists (He & Godfroid, 2019; Simpson-Vlach & Ellis, 2010). For example, He and Godfroid (2019) utilized teacher evaluations to assess the usefulness of academic words in the COCA and COCA Academic corpora.

In summary, the present study seeks to address a gap in research by integrating two perspectives: the receptive vocabulary knowledge of Saudi engineering students and the pedagogical evaluations of engineering lecturers. By doing so, it bridges the divide between corpus-based word lists and real-world ESP vocabulary learning contexts. While previous studies by Dang et al., (2022a, 2022b) have investigated learners' vocabulary knowledge and teacher evaluations, their focus was limited to general high-frequency vocabulary. While they provide valuable insights into learners' receptive knowledge and teacher assessments of general vocabulary, they do not address the specialised domain of engineering technical vocabulary. Therefore, the present study attempts to bridge this gap by investigating teachers' perceptions of the pedagogic usefulness of an engineering technical vocabulary list and Saudi engineering students' receptive knowledge of these vocabulary items.

6.2 Research Questions

This study aims to answer the following research questions:

1. To what extent do Saudi undergraduate engineering students receptively recognize engineering technical vocabulary items?
2. Which technical engineering vocabulary items do teachers perceive as being useful for pedagogical purposes?
3. What is the relationship between engineering teachers' perceptions of the usefulness of technical vocabulary and Saudi undergraduate engineering students' receptive knowledge of engineering technical vocabulary?

6.3 Methodology

This section describes the methodology of Study 3, beginning with a description of the participants of the study, followed by data collection methods with a detailed description of the three instruments used in the study: Updated Vocabulary Levels Test (UVLT) (Webb, et al., 2017), a Yes/No test of engineering technical vocabulary, and a teacher survey on the pedagogic usefulness of the engineering technical vocabulary list. Finally, it describes the analytic approaches employed in the study.

6.3.1 Participants

The study involved 78 final-year engineering students from Northern Borders University (NBU) in Saudi Arabia. Their ages ranged from 22 to 26 years. The participants represented four major engineering disciplines: Electrical, Civil, Mechanical, and Industrial Engineering. As illustrated in Table 6.1, the sample included 12 Electrical Engineering students, 34 Mechanical Engineering students, 19 Civil Engineering students, and 13 Industrial Engineering students.

Table 6-1: Student Participants Information

Participants	Field of Study	Study Year
12	Electrical Engineering	Final Year
34	Mechanical Engineering	Final Year
19	Civil Engineering	Final Year
13	Industrial Engineering	Final Year
Total = 78		

Table 6.2 provides detailed information about teacher participants, highlighting their disciplinary backgrounds, years of teaching experience, and academic ranks. The group consists of 20 teachers across various engineering fields, including Industrial, Mechanical, Mining, Chemical, and Electrical disciplines. Their experience ranges from 8 to 25 years, with most having over a decade of experience. The participants' academic ranks vary, with five professors at the highest level of seniority, a majority of associate professors, and one assistant professor. The participants were of different nationalities, including engineering teachers from Algeria, Egypt, Pakistan, Tunisia, and Saudi Arabia.

Table 6-2: Teachers' information

No	Discipline	Experience in Teaching	Academic Rank	Nationality
1	Industrial	25 years	Professor	Algerian
2	Industrial	20 years	Professor	Egyptian
3	Mechanical	23 years	Professor	Egyptian
4	Mining	23 years	Professor	Egyptian
5	Chemical	22 years	Professor	Pakistan
6	Mechanical	13 years	Associate Professor	Egyptian
7	Chemical	11 years	Assistant Professor	Tunisian
8	Industrial	14 years	Associate Professor	Tunisian
9	Electrical	8 years	Assistant Professor	Saudi
10	Mechanical	18 years	Associate Professor	Saudi
11	Mechanical	12 years	Associate Professor	Algerian
12	Industrial	14 years	Associate Professor	Egyptian
13	Electrical	12 years	Associate Professor	Egyptian
14	Mining	11 years	Associate Professor	Egyptian
15	Mining	14 years	Associate Professor	Pakistan
16	Mechanical	9 years	Associate Professor	Egyptian
17	Electrical	11 years	Associate Professor	Tunisian
18	Chemical	15 years	Associate Professor	Tunisian
19	Mechanical	13 years	Associate Professor	Saudi
20	Electrical	12 years	Associate Professor	Saudi

6.3.2 Data Collection

This section presents information on data collection. It begins with a detailed description of two instruments used for data collection from student participants: the UVLT and the Yes/No engineering vocabulary test. Next, it describes a five-point Likert scale survey instrument used to collect data from the teacher participants.

6.3.2.1 Updated Vocabulary Levels Test

The Updated Vocabulary Levels Test adopted from Webb, et al., (2017) was administered to assess the general receptive vocabulary levels of student participants in the current study. The test was originally developed by Nation (1983, 1990) and subsequently revised by Schmitt, Schmitt, and Clapham (2001).

The UVLT used a word-definition matching format. Each level of the test comprised ten sections, with each section containing six words and three definitions. Test-takers were required to select three words out of the six provided and match them with the three given definitions. In its original form, the VLT consisted of five levels: the 2,000-word level (2K), 3,000-word level (3K), 5,000-word level (5K), 10,000-word level (10K), and the AWL. The 2K and 3K levels are generally considered sufficient for effective communication in English. The 5K level represents vocabulary of moderate frequency. The 10K level targets less common vocabulary, while the AWL assesses knowledge of academic vocabulary.

Match the word with correct meaning

Sample question A (2k level).

Words	**Meaning**
1 blame	___________ *make*
2 elect	___________ *choose by voting*
3 jump	____________ *become like water*
4 threaten	
5 melt	
6 manufacture	

Figure 6-1:Format of the VLT (Nation, 1983, 1990; Schmitt, Schmitt, & Clapham, 2001)

In this study, the UVLT was chosen in his study because it has been validated and is widely used by educators and academics to evaluate learners' vocabulary levels in various contexts (e.g., Stæhr, 2008; Webb & Chang, 2012). This test aimed to assess the participants' general receptive knowledge of vocabulary to classify them based on their general proficiency in vocabulary into beginner, intermediate, and advanced groups.

The UVLT matching exercises evaluate learners' ability to identify words and accurately match them with their meanings. The correct answers for Sample Question A are Manufacture means 'make' (6), elect means 'choose by voting' (2), and melt means 'become like water' (5).

6.3.2.2 Yes/No Test of Engineering Technical Vocabulary Receptive Knowledge

A Yes/No test was designed to assess the students' receptive knowledge of ETV items developed in the previous study. This test was based on the format of Meara's (1992) Yes/No vocabulary test named English as a Foreign Language Vocabulary Test, which was developed to measure EFL learners' receptive knowledge of English vocabulary, containing real words and pseudowords. Pseudowords are words that conform to the language's phonological constraints but lack semantic content (Huibregtse et al., 2002). Yes/No tests are commonly used to measure receptive vocabulary knowledge (Anderson & Freebody, 1983; Harsch & Hartig, 2015, Meara & Buxton, 1987; Mochida & Harrington, 2006). The underlying assumption of this test, as with all tests of vocabulary breadth is that if a learner must first recognize the form of a word—its spelling and pronunciation—before they can understand its meaning and use it appropriately. This perspective underscores the idea that without initial word recognition, further engagement with it is unlikely (Meara, 1996b: Nation, 2001). The use of Yes/No tests appears to be an effective way to measure how much a learner of a foreign language knows about receptive vocabulary (Huibregtse et al., 2002).

In this study, a Yes/No test was preferred over a multiple-choice recognition test (e.g., Willis & Ohashi 2012) because it enables researchers to test a large number of target words with many participants (Culligan 2015; Huibregtse et al. 2002; Read, 2000; Schmitt et al., 2011). The Yes/No test format allows a higher sampling rate of items for reliable assessment because a large number of items can be tested in a limited period (Beeckmans, Eyckmans, Janssens, Dufranne, & Van de Velde, 2001; Harrington & Carey, 2009; Meara & Buxton, 1987; Mochida

& Harrington, 2006; Pellicer-Sánchez & Schmitt, 2012; Read, 2000). Additionally, the Yes/No test format is not difficult to construct, administer, or score (Harrington & Carey, 2009; Meara & Buxton, 1987; Mochida & Harrington, 2006). The next section presents a description of the format of the Yes/No vocabulary test used in this study.

6.3.2.2.1 Target Words

In this study, the target words consisted of the 930 ETV items developed in Study 2. These were categorized into high-, mid-, and low-frequency vocabulary bands based on Schmitt and Schmitt's (2014) framework, as outlined in Chapter 5 (see Table 5.4). For this study, samples of 210 words were selected from the 930 ETV list using stratified randomization for several reasons. Firstly, administering a test covering the entire 930 ETV items would be impractical, as participants might be unwilling to commit to such an extensive assessment. Secondly, the same set of real words was used for teacher evaluations of the ETV list to gain insights into the pedagogical value of these words. Additionally, evaluating the entire list would be time-consuming and burdensome for teachers. This approach ensured a manageable and effective assessment for both learners and educators.

To ensure balanced selection, the target words were evenly distributed across three Yes/No tests based on their frequency bands (high-, mid-, and low-). Test Form1 included 70 ETV items from the EMDC high-frequency band, Test Form 2 included 70 ETV items from the mid-frequency band, and Test Form 3 included 70 ETV items from the low-frequency band. Each test consisted of 70 real words and 30 pseudowords, resulting in a total of 210 real words and 90 pseudowords used in this study (see Appendix H). The real words were engineering technical vocabulary identified in the EMDC across BNC/COCA level lists plus supplementary lists. These words were selected based on their frequency, keyness, and specialised meaning relationships (see Section 5.2 for details).

6.3.2.2.2 Pseudowords

Pseudowords were used in the Yes/No test to control for guessing of answers by the participants (Nation & Webb, 2011). These are non-existent words that resemble real words in the language being tested, for example, '*porfume*' and '*freath*' (Meara & Buxton, 1987). Pseudowords were developed based on syllables of words from each frequency range of the target words and distributed randomly within the test (Huibregtse et al., 2002). All pseudowords were checked by native speakers to ensure that they followed the phonological rules of English and did not correspond to actual words (Appendix I). The main function of pseudowords is to limit the participants' tendency to guess. The inclusion of pseudowords in the study is based on the assumption that individuals who are familiar with all real words will respond positively to real words but negatively to pseudowords (Meara & Buxton, 1987). If participants indicate that they are familiar with pseudowords, their results are adjusted accordingly (Beeckmans et al., 2001; Read, 2000). Following the approach of Schmitt et al. (2011), data from six participants who endorsed more than 10% of the pseudowords were excluded.

TRAVERSE	[√] YES	[ ] NO
LOBE	[ ] YES	[√] NO
Tunne	[ ] YES	[√] NO

You know the meaning of the word "traverse" – in this case, you tick YES.

You don't know the word "lobe" – in this case, you tick NO.

Figure 6-2: Yes/No test of the ETV list

In the instructions, participants were asked to indicate whether or not they knew the technical meaning of each word by responding with ‘yes’ or ‘no’ (see Figure 6.2). They were informed that the tests included non-existent words (pseudowords), but no details were provided regarding their number or location in the tests (Huibregtse et al., 2002).

This is based on the assumption that if participants know all the words, they will tick ‘*Yes*’ to Any participants who checked more than 10% of the pseudowords in the test (3 out of 30 pseudowords) were excluded from the analysis, as recommended by Schmitt et al. (2011) and applied by Dang, et al., (2022a).

6.3.2.3 Teacher Survey of Evaluating Pedagogic Usefulness ETV List

Three surveys using a five-point Likert scale were designed to assess teachers’ perceptions of the pedagogic usefulness of the ETV list. The scale ranged from 1 (least useful) to 5 (very useful). The surveys contained the 210 ETV list, with the same words included as real words in the Yes/No tests. To facilitate analysis, all words were organized by their frequency levels, as presented in Table 6.3.

Table 6-3: Distribution of Survey Items

Survey	TV Real Words	Examples,
1	70 High-frequency TV	drill, converter, curves, spacer, filters, carbon, coolant, rays
2	70 Mid-frequency TV	Ceramic, turbine, torque, electromagnetic, armchair, Vacuum, condenser, pores
3	70 Low-frequency TV	Microfiltration, modularized, nonadaptation, convection, exponential, speckle, axial, ergonomic
Total	210 TV (Real words)	

As shown in Table 6.3, the survey was divided into three sections, each containing 70 ETV items. The first survey contained 70 ETV items from the high-frequency level. The second survey consisted of 70 ETV items from the mid-frequency level. The third survey contained

70 engineering technical vocabulary from the low-frequency level. The total number of ETV is 210 (real words). This format was informed by a pilot study conducted with three engineering faculty members from NBU. These lecturers each had over ten years of experience in engineering. They completed a 210-item survey and unanimously suggested that dividing the items into three sections by frequency bands would reduce fatigue and improve focus compared to completing a single, lengthy survey. The full test is included in the Appendix H.

In this study, the Likert rating scale format was selected due to its effectiveness in eliciting participants' opinions regarding various aspects of language learning (Brown & Rodgers, 2002). This survey format offers detailed insights into the perceived usefulness of the target words. One drawback of this format is that it requires more time to complete due to its task demands. However, since teachers had the flexibility to complete the surveys at their convenience, conducting the scale surveys remained feasible. A five-point scale was utilized because, in comparison to a seven- or 10-point scale, it reduces confusion among participants regarding the different points on the scale, thereby minimizing the likelihood of unreliable responses (Dörnyei & Taguchi, 2010). Additionally, a five-point scale allows for a more comprehensive exploration of participants' perspectives compared to a three-point scale. A scale with too few points may not yield sufficient information for researchers (Dörnyei & Taguchi, 2010).

Participants were instructed to rate the usefulness of the ETV list using a five-point scale: 1. Least useful, 2. Slightly useful, 3. Moderately useful, 4. Useful and 5. Very useful (see Figure 6.3). The surveys were distributed via email as attached Excel files, allowing teachers to complete them at their convenience and in alignment with their work schedules. Teachers downloaded the surveys, filled them out, and returned them to the researcher once completed. Following submission, the researcher reviewed the responses and, when necessary, requested additional clarification or information from the participants. The rationale for using Excel aligns with Dang's (2017) argument that, compared to paper-and-pencil or online formats, employing an attached Excel format provides a more efficient and effective means of data collection from teachers. Unlike the paper-and-pencil format, emailing surveys as Excel files allows researchers to gather data from geographically dispersed teacher populations with minimal disruption to their busy schedules (Dörnyei & Taguchi, 2010). Moreover, in contrast to online surveys, using Excel files results in higher response rates and more valid data. If all ETV items had been included in a single online survey, it might have discouraged teacher participation or led to fatigue. Distributing the items into three shorter surveys (each containing 70 words, as previously mentioned) addressed this issue, allowing teachers to complete them at their convenience. Figure 6.3 below describes the instructions involved in teacher surveys of rating word usefulness.

In the following section, please respond to the question: **To what extent is the word useful for your students? (For engineering purposes, not general purposes).** Use the following scale to rate your perception of the usefulness of each word: 1. Least useful 2. Slightly useful 3. Moderately useful 4. Useful 5. Very useful.

Example Response: *If you believe that the word* Spacer *is slightly useful,* Converter *is useful, and* Carbon *is very useful, your responses would be:* **Converter***: 4,* **Spacer***: 3* **Carbon**: 5

No	Word	Degree of usefulness
1	**Converter**	4
2	**Spacer**	3
3	**Carbon**	5

Figure 6-3: Instructions for teacher surveys of word usefulness

6.3.3 Analytic Approach

This section outlines the data analysis procedures applied to the three sets of data collected in this study. It begins with the analysis of student data from the UVLT, followed by the examination of data obtained through the Yes/No tests. Finally, it presents the analysis of data gathered from the teacher participants' surveys.

6.3.3.1 Analysis of the Updated Vocabulary Level Test (UVLT)

The data collected from 78 students who took vocabulary-level tests were used for the analysis. Their performance in the UVLT was marked and scored, with participants receiving one point for each correct response, with a maximum possible score of 30 per level. SPSS version 27.0 was used to compute the mean and standard deviation for each level. To examine differences across levels, a one-way repeated measures ANOVA was conducted, and post-hoc tests were applied where significant differences were found to identify the specific levels of difference. Following Schmitt, Schmitt, and Clapham's (2001) guidelines (Xing & Fulcher, 2007), a mastery threshold of 26 out of 30 (87%) was set at each level.

The main purpose of the UVLT test in the present study was to group the student participants according to their general vocabulary levels of proficiency into three groups. The beginner group consisted of learners who had yet to master the 2K level, with a score of less than 26 out of 30 (87%) at this level. The intermediate group comprised learners who had yet to master the 3K level, meaning those who scored 26 out of 30 or 87%. Finally, the advanced group consisted of 26 participants who had at least mastered the 3K vocabulary level. This suggests that these learners have a good command of high-frequency words (1K to 3K) but might struggle with lower-frequency vocabulary (Laufer & Nation, 1999). When interpreting the UVLT scores, the focus is on individual-level scores rather than the total, as these indicate specific areas for targeted vocabulary learning (Webb et. al, 2017). For detailed results, see Section 6.4.1 (Table 6.4 and Table 6.5).

6.3.3.2 Scoring the Yes/No Vocabulary Test

Various scoring methods exist for Yes/No receptive tests, including the calculating hits method, hits minus false alarms (h-f) method, correction for guessing (Cfg) method, Δm Meara's formula, and Index Signal Detection Theory (I_{SDT}). However, to evaluate which Yes/No test scoring method is more effective in distinguishing word knowledge at different frequency levels, it is essential to consider how each scoring method handles the balance between correct identification (hits) and overestimation (false alarms). Each method reveals unique insights into participants' receptive vocabulary knowledge at high-, mid-, and low-frequency word levels. The current study employs the I_{SDT} formula to analyse participants' responses to the Yes/No tests, assessing their receptive knowledge of the ETV list. The I_{SDT} formula has proven more effective in addressing individual response biases (Beeckmans et al., 2001; Huibregtse et al., 2002; Mochida & Harrington, 2006). Other scoring methods were also reviewed and synthesized to support the selection of the I_{SDT} formula (see Table 6.8).

Moreover, scoring the Yes/No test by simply calculating the hits and correct rejections as equivalent responses is problematic (Huibregtse et al., 2002). It can be challenging to interpret a basic score of all correct responses (Mochida, & Harrington, 2006). Another method of scoring is to calculate the proportion of hits minus the false alarm rate (*h-f*). False alarms are central to the scoring and interpretation of test performance (Mochida & Harrington, 2006). Although this strategy accounts for false alarms, it could overestimate vocabulary knowledge when the false alarm rate is low (Huibregtse et al., 2002, p. 231). According to Huibregtse et al. (2002), very low false alarm rates point to overly cautious performance that results in scores that understate actual knowledge whereas large false alarm rates point to a significant degree of guessing and scores (whether corrected or not) of low reliability. For less severe performance, a given false alarm rate may provide a varied corrective effect depending on the relative hit rate and the type of correction formula utilized (Beeckmans et al., 2001; Huibregtse et al., 2002; Eyckmans, 2004). Other approaches were adopted to address guessing in scoring Yes/No responses "that take into account individual differences in the proportion of hits and false alarms produced" (Mochida & Harrington, 2006, p. 76). These include the correction for guessing (Cfg) procedure (Anderson & Freebody, 1983; Cameron, 2002; Meara, 1989; Meara & Buxton, 1987), Meara's (1992) formula, and I_{SDT} Theory (Huibregtse et al., 2002), as described in Figure 6.4 below.

1) Correction for guessing (*cfg*):

$$P(h) = \frac{(h)-(f)}{1-(f)}$$

where *P(h)* = true hit rate, *h* = observed hit rate, *f* = observed false alarm rate.

2) Meara' s Δm

$$\Delta \mathrm{m} = \frac{(h-f)}{(1-f)} - \frac{f}{h}$$

3) Index of Signal Detection (I_{SDT})

$$I_{SDT} = 1 - \frac{4h(1-f)-2(h-f)(1+h-f)}{4h(1-f)-(h-f)(1+h-f)}$$

Figure 6-4: Three methods of correction for guessing in Huibregtse et al. (2002)

The *Cfg* formula calculates a score as the proportion of *h-f* divided by the proportion of pseudowords correctly rejected (Mochida & Harrington, 2006). It is based on a 'blind guessing model' which assumes that the respondent is either aware of the word or making a random guess; in other words, L2 vocabulary knowledge is categorical (Huibregtse et al., 2002, p. 231). The shortcoming of this method is that "it stresses the hit rate over the false alarm rate" (Mochida & Harrington, 2006, p. 77). To overcome this shortcoming, Meara (1992) suggested a second formula. However, this formula proved to be overly conservative, as it produced scores that "over-corrected for false alarms in general and in particular yielded uninterpretable" results (Huibregtse et al., 2002, p. 245). Both Cfg and Meara formulas have overlooked an individual response bias.

Word	Pseudoword
Hit	False alarm
Miss	Correct rejection

Figure 6.5: Response alternatives matrix for Yes/No test items

When calculating scores on a Yes/No test, it is important to consider four types of responses: two correct and two incorrect. Additionally, participants' tendencies to guess and their various response styles should be taken into account, as illustrated in Figure 6.5. The two types of correct responses are 'yes' for a real word (hit) and 'no' for the correct rejection of a pseudoword. The incorrect responses are 'no' to a real word (miss) and 'yes' for a pseudoword (false alarm) (Huibregtse et al., 2002). In contrast to correct rejection rates, hit rates are thought to represent underlying vocabulary knowledge (Mochida & Harrington, 2006). Of primary importance is the number of hits, adjusted for the false alarm rate (Mochida & Harrington, 2006), even if both hits and the correct rejections of pseudowords are valid responses.

6.3.3.3 Validity of the Yes/No Test

The Yes/No test is widely recognized as a valid measure of receptive vocabulary knowledge in previous studies (e.g., Dang et.al, 2022; Meara, 1996; Read, 2000) demonstrating a strong correlation with other test formats such as multiple-choice, matching, and cloze tests. In this study, the Yes/No test was used to assess students' receptive knowledge of the ETV list developed in Study 2. According to Messick (1989), validating a test requires both logical reasoning and empirical evidence derived from quantitative and qualitative data. Schmitt, Schmitt, and Clapham (2001) applied item analysis, profile of section analysis, and factor analysis to validate different versions of the Vocabulary Levels Test administered to 801 English learners. Similarly, this study employed three validity analyses: expert-based face and content validity, item analysis, and profile of the analysis of the sections, ensuring a comprehensive evaluation of the test's validity.

6.3.3.3.1 Face and Content Validity

To ensure the face and content validity of the ETV Yes/No test, the instrument was exclusively designed using target vocabulary from the 930 ETV items developed in the previous chapter. Face and content validity are grounded in expert evaluation of content relevance and representativeness (Neuendorf, 2016), relying on expert judgement rather than statistical measurement (Gay, Mills & Airasian, 2006). Two linguistics experts, both English professors at NBU in Saudi Arabia, reviewed the test instruments developed for this study. Their insights were particularly valuable in refining the selection and development of pseudowords, enhancing the overall validity of the tests.

6.3.3.3.2 Item Analysis

In this study, item analysis was performed by determining the facility index (IF), which provided the discrimination indices (point-biserial) for each item. Additionally, the percentage of test-takers who answered each question correctly was calculated and is presented in Table 6.4.

Table 6-4: Items Analysis of Yes/No Tests

		Yes/No	
Participants		Facility Value	Discrimination Indices
N	Valid	50	50
	Missing	0	0
Mean		.84	.70
Median		.88	.75
Std. Deviation		.15	.30
Minimum		.35	-.30
Maximum		1.00	1.00

Table 6.4 provides a statistical summary of the Yes/No vocabulary test taken by 50 participants (pilot), with key indicators analysed for each item's performance. The mean score for the items was .84, reflecting that, on average, most items were relatively easy, with 84% of participants answering correctly. Most of the words indicated unknown by the students were from the low-frequency level (e.g., *zircaloy, foulant, speckle, Poiseuille*). The slightly higher median (0.75) suggests that at least half of the items have a high discrimination level, enhancing the test's ability to assess true knowledge differences. The standard deviation (0.30) indicates larger variability in discrimination indices compared to facility values, reflecting greater differences in how well individual items distinguish between strong and weak performers. The minimum and maximum values (-0.30 to 1.00) show that some items may not discriminate well (or may even favour lower-performing students) with a minimum of -0.30, while others are highly effective at distinguishing knowledge levels, achieving a perfect 1.00. Consistent with Schmitt, Schmitt, and Clapham's (2001) argument, it is affirmed that the discrimination indices for the test are acceptable. This takes into account the nature of vocabulary acquisition, which occurs as individual units. It is common for less proficient learners to know some relatively low-frequency words, while more proficient learners may still show gaps in their knowledge of higher-frequency vocabulary.

6.3.3.3.3 Profile Analysis of the Sections

Profile analysis of the sections is another method for estimating the validity of a level test. This approach involves comparing participants' performance across different sections of the test (Schmitt et. al, 2001). Research has consistently shown that learners generally acquire higher-frequency words before mastering lower-frequency ones (e.g., Nation, 1990). Consequently, the validity of the Levels Test can be partially assessed by determining whether learners perform better on higher-frequency sections than on lower-frequency sections.

In this study, a profile analysis was applied to the three sections of the Yes/No test: high-frequency, mid-frequency, and low-frequency levels. This analysis aimed to evaluate participants' performance across these frequency levels. The results of the estimated performances for each section are presented in Table 6.5 below.

Table 6-5: Profile Analysis of the Items in Yes/No Tests

		Yes/No		
		High-frequency	Mid-frequency	Low-frequency
N	Valid	50	50	50
	Missing	0	0	0
Mean		92.53	85.50	75.55
Minimum		74	51	43
Maximum		100	100	91

Table 6.5 shows that the mean scores across the three frequency levels were 92.53 for high-frequency words, 85.50 for mid-frequency words, and 75.55 for low-frequency words. These results indicate that students demonstrated the highest mastery of high-frequency words, followed by mid-frequency words, with low-frequency words showing the lowest levels of correct responses. This finding aligns with the findings of Schmitt, Schmitt, and Clapham's (2001) results, who reported mean scores (out of 30 possible) of 25.29 (SD = 5.80) for the

2,000 level, 21.39 (SD = 7.17) for the 3,000 level, 18.66 (SD = 7.79) for the 5,000 level, and 9.34 (SD = 7.01) for the 10,000 level. Similarly, Read (1988) observed comparable differences between frequency levels and conducted implicational scaling, which found the sections to be 'highly scalable', with coefficients of scalability at .90 and .84 across two administrations. This scalability indicates that achieving mastery at a lower frequency level implies mastery of all higher frequency levels. This study adopted a criterion of mastery defined as achieving 26 out of 30 possible correct responses per level. This criterion is based on Guttman scalability analysis (Hatch & Lazaraton, 1991), as recommended by Read (1988) and Schmitt, Schmitt, and Clapham (2001).

6.3.3.4 Reliability of the Yes/No Test

The reliability of the Yes/No test instruments was assessed using Cronbach's alpha, a widely recognized measure of internal consistency (Meara & Buxton, 1987). Cronbach's alpha was calculated for each frequency band – high-, mid-, and low- – within the Yes/No ETV test. The results demonstrated high reliability across all bands, with values of .91 for high-frequency items, .93 for mid-frequency items, and .89 for low-frequency items. These indices exceed the commonly accepted threshold of 0.7, as recommended by Nunnally and Bernstein (1994), indicating strong internal consistency in item responses.

This high level of reliability confirms that the Yes/No test items are dependable and consistent tools for measuring vocabulary knowledge across different word frequency levels, affirming the test's appropriateness for assessing receptive vocabulary knowledge. These findings align with those of Read (1988) and Schmitt, Schmitt, and Clapham (2001), who reported Cronbach's alpha values of .91 and above in their respective studies of Level Tests, further validating the reliability of the Yes/No test format.

6.3.3.5 Bootstrapping Descriptive Statistics of Yes/No ETV Test

Bootstrapped descriptive statistics, including means, standard deviations, and confidence intervals (CIs), were utilized to evaluate participants' receptive knowledge of engineering technical vocabulary across three learner groups: beginners, intermediate, and advanced. Bootstrapping, a robust statistical technique, simulates study replication by resampling from the population (Beasley & Rogers, 2009; LaFlair, Egbert, & Plonsky, 2015) (see Table 6.9). CIs offer valuable insights by indicating whether the difference between two mean scores is statistically significant and stable. This is determined by checking whether the mean of one group falls within or outside the CI of another group's mean (LaFlair et al., 2015). CIs represent a range of values around an observed mean score that likely (at a given probability level, typically 95%) contains the true population mean (LaFlair et al., 2015).

While there are various ways to interpret CIs, their primary function is to contextualize mean scores within the range of possible values that might reflect the true population mean, rather than the sample mean (Cumming, 2012). As Carl Sagan (1996) aptly stated, CIs serve as 'a quiet but insistent reminder that no knowledge is complete or perfect' (pp. 7–28). Similar to standard deviations, considering CIs whether numerically or visually helps mitigate the tendency to view sample data and their mean scores as absolute (LaFlair et. al, 2015)

6.3.3.6 Data Analysis of the Teacher Survey

The analysis of data from teacher participants was conducted in two stages. In the first stage, bootstrapped descriptive statistics were used to calculate means, standard deviations, and confidence intervals (CIs) for the data collected through teacher surveys on the pedagogic usefulness of the engineering vocabulary list, as detailed in Section 6.4.4. The analysis evaluated teachers' assessments of 210 target words categorized into high-, mid-, and low-frequency bands, following Schmitt and Schmitt's (2014) framework (see Table 6.10). The items were ranked based on their mean scores and standard deviations as rated by the teachers. This ranking helped determine the pedagogic usefulness or relative importance of each word in the list, as perceived by experienced engineering educators (see Appendix G).

In the second stage, bootstrapped Pearson product-moment correlations were performed to explore potential relationships between learner vocabulary knowledge and teacher perceptions of the pedagogic usefulness of the ETV list. Bootstrapping was employed for its robustness, as it treats the current sample as the population, repeatedly draws random samples from it, and conducts the statistical test on these samples (Plonsky, 2015). This approach aligns with He and Godfroid's (2019) study, which examined the relationship between frequency and teacher perceptions of academic vocabulary usefulness. The interpretation of correlation values (r) between 0 and 1 adhered to Plonsky and Oswald's (2014) benchmarks for effect sizes in L2 research, where r values around .25 indicate a small effect, .40 a medium effect, and .60 a large effect (p. 889).

6.4 Results and Discussion

This section presents results and discussion to address the research questions in Study 3. It begins by reporting participants' general vocabulary proficiency levels, assessed using the UVLT, a standardized instrument for measuring receptive vocabulary knowledge. Participants were categorized based on their vocabulary proficiency levels. The next section details the results of the analysis of data collected through the Yes/No ETV tests administered to student participants, which evaluated their receptive knowledge of engineering-specific technical vocabulary. Finally, the results of the teacher surveys are presented, focusing on their perceptions of the pedagogical usefulness of the engineering technical vocabulary items.

6.4.1 Saudi Engineering Students' Vocabulary Knowledge

This section presents results from the UVLT administered to 78 participants to assess their receptive knowledge at the 2,000-word (2K), 3,000-word (3K), and 5,000-word (5K) frequency levels.

Table 6-6: Learners' VLT Mean Scores (N = 78)

Frequency Levels	Correct responses			
	Mean	Std. Deviation	Minimum	Maximum
2K	18.22	8.53	3	30
3K	13.01	8.739	2	28
5K	11.01	8.508	2	30
Total	42.24	22.777	7	88

Table 6.6 provides an overview of the participants' scores on the UVLT. The results reveal that participants' receptive vocabulary knowledge decreases as word frequency decreases. The high standard deviation (SD = 22.77) reflects considerable variation in the participants' receptive knowledge across word frequency levels. The mean scores for the 2K, 3K, and 5K word levels were 18.22, 13.01, and 11.01, respectively. This decline in performance as frequency levels decrease aligns with established research in second language acquisition, which suggests that learners tend to acquire and retain high-frequency words more easily than low-frequency words due to their greater exposure to these words in everyday language use (Laufer & Nation, 1999; Webb & Chang, 2012). The overall mean of the total score of participants was 42.24, with a minimum score of 7 and a maximum score of 88.

The results of the one-way repeated measures ANOVA revealed a significant difference in correct responses among the three frequency levels (2K, 3K, and 5K). The F-value was 5.63, with a P-value less than .001, indicating a medium to large effect size ($\eta^2 = .36$). This suggests that the frequency level has a substantial impact on vocabulary knowledge. The P-value of less than 0.05 further confirms the presence of statistically significant differences in the mean correct responses across the 2K, 3K, and 5K frequency levels.

This finding reflects the general receptive vocabulary levels of Saudi engineering undergraduate students, aligning with the typical lexical profile observed in EFL contexts. Students tend to have stronger knowledge of higher-frequency words compared to mid- and low-frequency words, indicating limited exposure to and mastery of less common vocabulary. This result supports findings from Schmitt et al. (2001) and Webb and Chang (2012), which highlight L2 learners' strong grasp of the 2K level words while struggling with less frequent vocabulary. In addition, Nation (2001) explains that L2 learners typically prioritize acquiring high-frequency words essential for basic comprehension before gradually progressing to mid- and low-frequency words, which are often associated with advanced or specialised contexts. Similarly, Dang (2017) observed that learners' knowledge of 2K-level words was significantly stronger than that of 3K- and 5K-level words among 275 participants assessed using the UVLT, reinforcing this progressive acquisition pattern.

Table 6-7: Vocabulary Levels of Each Learner Group (N=78)

Groups of learners (vocabulary level)	**Number of learners**	**Vocabulary level**
Beginners	19	Yet to master the 2K level **(Scored less than 26/30 (87%) at the 2K level)**
Intermediate	33	Mastered the 2K level **(Scored 26/30 (87%) or higher at the 2K level)**
Advanced	26	Mastered at least the 3k level (Scored 26/30 (87%) or higher at 3k level)

Table 6.7 shows that the participants are categorized into three groups labelled as beginners, intermediate, and advanced learners (Schmitt et. al, 2001). The first group comprises 19

beginner students who have yet to master the 2K level, scoring less than 26 out of 30 (87%) at this level. This implies that these learners have a very limited vocabulary, lacking knowledge of even the most basic and frequent words in English (Nation & Waring, 1997). The second group (intermediate) includes 33 participants who had yet to master the 3K level. These participants scored 26 out of 30 or 87%. The result indicates that they have a fair understanding of the most frequent words (2K) but have not yet acquired a substantial number of words at the 3K level and beyond (Webb & Chang, 2012).

Finally, the advanced group consists of 26 participants, who had mastered at least the 3K vocabulary level. This suggests that these learners have a good command of high-frequency words (1K to 3K) but may struggle with lower-frequency vocabulary (Laufer & Nation, 1999). These findings highlight the diverse vocabulary proficiency levels among the learners and underscore the need for differentiated vocabulary instruction targeting each group's specific needs (Nation, 1990). The mastery threshold for the UVLT has varied across research, with earlier scholars like Schmitt, et al. (2001) recommending a score of 26 out of 30 (87%) to indicate true mastery at a level. Interpretation of UVLT results focuses on individual-level scores rather than a combined score, as higher-frequency words are generally more valuable for foundational language use. In the original VLT guidelines, Nation (1983) suggested that scores below 66% (20 out of 30) at any level indicate a need for further study of those words.

6.4.2 Saudi Engineering Students' Receptive Knowledge of Engineering Technical Vocabulary

This section presents the results of the Yes/No tests assessing Saudi engineering undergraduate students' receptive knowledge of ETV. It examines students' receptive vocabulary knowledge across three frequency levels (high, mid, and low) and provides an overall summary of their receptive knowledge. Additionally, it highlights differences in ETV knowledge according to students' general vocabulary level size (grouped into beginners, intermediate, and advanced).

The analysis was conducted using the Index of Signal Detection Theory (I_{SDT}) correction method for the participants' responses to ETV Yes/No tests. Table 6 summarizes the mean scores (M) and standard deviations (SD) for the three frequency levels: high, mid, and low.

Table 6-8: Participants' Receptive Knowledge of ETV Based on the ISDT Scoring Method

Frequency Levels	M	SD
High-frequency	5.09	5.11
Mid-frequency	4.32	5.87
Low-frequency	5.33	6.76
Overall	4.91	5.97

Table 6.8 reveals that Saudi engineering students' receptive knowledge of ETV, as measured by the I_{SDT} method, demonstrates limited overall mastery (M = 4.91, SD = 5.97). The high standard deviation suggests significant variability in individual performance, with some learners scoring markedly higher or lower than the group average. Notably, while high-frequency words yielded marginally better recognition (M = 5.33, SD = 6.76), the scores for low-frequency terms were unexpectedly comparable (M = 5.09, SD = 5.11), with mid-frequency words trailing slightly (M = 4.32, SD = 5.87). This near-parity across frequency bands contrasts with established vocabulary acquisition theories (Nation, 2001; Schmitt et al., 2001), which posit a clear hierarchy favoring high-frequency word retention. The I_{SDT} method's consistency evidenced by stable SDs may reflect its effectiveness in controlling for guesswork (Eyckmans, 2004), but its flattening of frequency-based distinctions could obscure the progressive learning curve described by Read (1988).

The students' stronger performance on high-frequency words aligns with exposure-based models (Beeckmans et al., 2001), yet the atypical elevation in low-frequency scores warrants caution. This pattern might indicate random guessing or task-specific artifacts, as low-frequency terms typically challenge L2 learners (Meara & Buxton, 1987). The I_{SDT} 's reliability for cross-group comparisons is evident, but its insensitivity to frequency effects underscores the need for triangulation with other measures (e.g., lexical decision tasks) to distinguish genuine knowledge from methodological noise. These findings echo Meara's (1996) critique of single-method assessments while affirming I_{SDT} 's utility as a baseline tool for bias-adjusted vocabulary estimation.

The I_{SDT} method was selected over traditional accuracy measures due to its ability to control for response bias and random guessing, which are particularly problematic in yes/no vocabulary tests (Eyckmans, 2004). This method provides more reliable estimates of actual vocabulary knowledge by accounting for participants' tendency to respond 'yes' regardless of actual word knowledge.

Table 6-9: Participants' Receptive Knowledge of ETV Based on Other Scoring Methods

Frequency	**Yes/No test scoring methods**							
Levels	Hits		h-		Cfg		Δm	
	M	SD	M	SD	M	SD	M	SD
High-frequency	51.31	12.88	51.57	21.15	-0.56	18.71	-0.98	0.49
Mid-frequency	51.31	12.88	38.68	29.30	4.23	13.24	-0.58	0.79
Low-frequency	37.03	18.96	32.04	36.11	2.77	9.21	0.050	2.37
Overall	143.97	36.23	-12.05	72.47	-0.08	0.77	-0.09	0.76

Table 6.9 presents the results of Yes/No receptive tests of engineering technical vocabulary items using different scoring methods: hits, hits minus false alarms (h-f), correction for guessing (Cfg), and Δm (Meara's formula), which were used to arrive at I_{SDT} Theory. The data in Table 6.9 indicates that the hits method has the highest overall mean (143.97, SD = 36.23), suggesting that participants are generally confident in recognizing engineering technical vocabulary. However, this score could be influenced by overestimation, as hits do not account for false positives. The variation across the frequency levels suggests that participants demonstrate firmer receptive knowledge of high-frequency technical vocabulary compared to low-frequency levels. The 'hits' method shows very similar mean scores for both high-frequency (51.31) and mid-frequency (51.31) words, but it decreases notably for low-frequency

words (37.03). This suggests that participants were better at identifying high- and mid-frequency words, but their performance dropped for less familiar, low-frequency words. The standard deviation remains relatively low (12.88 for high- and mid-frequency, 18.96 for low-frequency), indicating consistent results across participants, though performance varies more with low-frequency words. This trend aligns with established findings in vocabulary research, which indicate that learners are generally more accurate in recognizing high-frequency words due to greater exposure (Laufer & Nation, 1999).

The results in Table 6.8 highlight a gradual decline in mean scores as word frequency decreases, underscoring the challenge learners face in recognizing mid- and low-frequency technical terms. While this method reflects word recognition, it does not account for guessing or overestimation, leading to a less refined measurement of true receptive knowledge. While simple, it is potentially inflated by false positives. This aligns with Mochida and Harrington's (2006) argument that relying on hit scores alone overlooks significant differences in guessing behaviour, as shown by variations in false alarms. Therefore, a correction for guessing is necessary. One approach is to calculate the hit rate minus the false alarm rate (H-F).

The hits-minus-false alarms (*H-F*) method shows an overall mean of -12.05 and a high standard deviation (SD = 72.47), indicating some sensitivity to word frequency while adjusting for overestimation by accounting for false alarms. The high inconsistency in this method demonstrates that while it captures differences in participants' knowledge across frequency levels, there is considerable variability, particularly at the lower frequency level. The data show a decrease in mean scores from high to low frequency (51.57 for high, 38.68 for mid, and 32.04 for low frequency), indicating that this method is sensitive to frequency variations, where the participants' ability to correctly identify words reduces as word frequency decreases. However, the substantial variability at mid- and low-frequency levels suggests participants' inconsistent knowledge of less common words. The limitation of this method is that it is more sensitive to guessing behaviour and tends to provide a more accurate reflection of a participant's actual vocabulary knowledge by adjusting for overestimation (Eyckmans, 2004). It is effective in controlling for false positives, making it superior to the raw 'hits method' in terms of providing a more accurate assessment. This result supports Huibregtse et al.'s (2002: 231) claim that, while this method does account for false alarm performance, it may still underestimate actual vocabulary knowledge when the false alarm rate is low.

Table 6.9 illustrates an overall low mean score (-0.08, SD = 0.77) from the correction for guessing (Cfg) method, implying heavy penalization for guessing and leading to a conservative evaluation of vocabulary knowledge. This approach may underestimate participants' true receptive knowledge, particularly at higher frequencies, by reducing the influence of inflated scores from random guessing. The Cfg method shows small mean scores across all frequency levels, with negative values at high frequency and positive values at mid- and low-frequency levels. This method yields small and mixed scores. For high-frequency words, it produces a slightly negative score (-0.56), implying that participants may have guessed incorrectly. Positive scores for mid-frequency (4.23) and low-frequency words (2.77) suggest that guessing was somewhat more accurate at these levels. The lower standard deviations across frequency levels (e.g., 9.21 for low-frequency) indicate less variability in results. The lower variability in scores compared to hits and h-f suggests that Cfg minimizes random guessing, a key issue in Yes/No vocabulary assessments (Mochida & Harrington, 2006). However, the method's low mean values may indicate an overly conservative approach, which could lead to underestimating learners' receptive vocabulary knowledge. As such, the resulting scores may not fully capture knowledge. The main limitation here is the emphasis on the hit rate over the false alarm rate (Mochida and Harrington, 2006). This formula proved overly conservative, often over-correcting for false alarms and, in cases where hits were relatively low and false alarms relatively high, producing uninterpretable or null scores (Huibregtse et al., 2002: 245). To address this issue, Meara (1992) proposed an alternative formula, as noted by Huibregtse et al. (2002: 230).

Table 6.9 indicates that Δm (Meara's Formula) method also provides a conservative estimate, with an overall mean score of -0.09 (SD = 0.76). Similar to Cfg, Δm adjusts for overestimation but may also lead to underestimated knowledge scores, especially at high-frequency levels. This scoring method yields consistently low mean scores, particularly at high (-0.98) and mid (-0.58) frequencies, with only a slight positive mean (0.050) at the low-frequency level. Δm's stringent scoring reduces score inflation, and the low standard deviations suggest minimal variation, indicating consistency across frequency levels. However, this conservative approach may underestimate participants' vocabulary knowledge, particularly at higher frequencies. This suggests that the method might be overly sensitive to overestimation, as participants are penalized heavily for false alarms, leading to negative scores. This aligns with Meara's (1989) observation that Δm provides a stricter measure, potentially reducing the inflation of scores caused by false positives. Therefore, Δm's harsh penalization makes it less effective at differentiating actual knowledge, particularly for high-frequency words, where participants are likely more accurate. The negative scores could imply that the method is too harsh in some cases, reducing its practical usefulness.

In summary, when compared with the results of the I_{SDT} method, the h-f method proved to be more effective in assessing participants' receptive vocabulary knowledge across different frequency levels. These findings align with previous studies, such as Beeckmans et al. (2001) and Eyckmans (2004), which emphasize the significance of correcting for false alarms to obtain more reliable evaluations of L2 receptive vocabulary knowledge. Moreover, the consistent performance of the I_{SDT} method across frequency levels indicates its potential as a reliable measure of general receptive knowledge, without being unduly influenced by word frequency. The next section further reports the results of receptive knowledge of engineering technical vocabulary across the three groups of learners.

6.4.3 Receptive Knowledge of Engineering Technical Vocabulary Across Groups of Learners

This section presents the variations in receptive knowledge of ETV according to students' vocabulary level size groups, using bootstrapped descriptive statistics to deepen our understanding of learners' familiarity with technical vocabulary. Students' general vocabulary level size was determined and grouped according to their results on the UVLT. As presented earlier, the participants were divided into three groups (19 beginners, 33 intermediate, and 26 advanced) based on their scores on Webb et al.'s (2017) UVLT. The scores of the participants in the Yes/No tests are further reported according to the students' group of general vocabulary size. The mean scores and confidence intervals indicate variations in vocabulary knowledge across these groups, suggesting that proficiency level is a significant factor influencing students' receptive understanding of technical terms in engineering.

Table 6-10: Bootstrapped Descriptive Statistics of Students' Receptive Knowledge of Engineering Technical Vocabulary Based on Proficiency Levels

Group of Learners (Vocabulary Levels)		**BCa 95% CI of mean**		
The proportion of students knowing the words	Mean	Lower	Upper	SD
Beginners (N = 19)	50.32	45.21	55.26	11.959
Intermediate (N = 33)	53.33	12.53	49.03	57.36
Advanced (N = 26)	62.81	10.084	59.46	66.42
All three groups combined (N=78)	55.76	12.582	52.90	58.43

Abbreviations: BCa, bias-corrected and accelerated; CI, confidence interval

Table 6.10 provides insights into students' receptive knowledge of ETV, categorized into three groups of learners' general vocabulary level size: beginners, intermediate, and advanced. The bootstrapped data reveals distinct trends in vocabulary knowledge across these groups. Beginners (N = 19) demonstrated a mean knowledge of 50.3% of the target words, with a 95%

confidence interval (CI) ranging from 45.2% to 55.3%. The standard deviation of 12.0 reflects variability within this group, possibly due to differing educational backgrounds or exposure to technical terms. This result indicates that beginners, while having some foundational vocabulary knowledge, generally recognize just over half of the technical vocabulary items tested. This finding aligns with previous studies that highlight the challenges faced by early language learners in mastering specialised vocabulary. Nation (2001) emphasizes that learners typically begin by acquiring high-frequency words and gradually expand their vocabulary to include more specialised, domain-specific terms. The low mean score observed in the beginner group suggests that these students may require additional training or exposure to engineering-specific vocabulary to build a stronger foundation in the subject matter. This observation supports the notion that targeted instruction in technical vocabulary is essential for beginners to bridge the gap between general vocabulary knowledge and discipline-specific terms (Read, 2000; Webb & Chang, 2012).

The bootstrapped analysis represented in the table shows that the intermediate group (N = 33) has a slightly higher mean knowledge level at 53.3%, though their reported CI of 12.5% to 49.0% appears unusually narrow, suggesting a possible reporting error. This score reflects modest improvement over the beginners, with an average receptive knowledge of slightly more than half of the target technical vocabulary items tested. Despite this gain, the relatively high standard deviation of 12.5 shows considerable variability in vocabulary knowledge within the intermediate group, potentially due to the mixed backgrounds and experiences of learners at this proficiency level.

The advanced group had the highest mean score at 62.8% (SD = 10.1), with a 95% confidence interval from 59.5% to 66.4%. This score indicates a stronger grasp of technical vocabulary among advanced students compared to beginners and intermediates. The narrower confidence interval and slightly lower standard deviation for this group suggest greater consistency in vocabulary knowledge. This result aligns with research suggesting that higher proficiency correlates with better technical vocabulary retention, as advanced students are more likely to have extensive exposure to technical language through study or practical application.

Finally, when combining all three groups, the overall mean score for receptive knowledge of engineering vocabulary is 55.8% (SD = 12.6), with a confidence interval of 52.9% to 58.4%. This overall mean reflects a moderate understanding of ETV across the samples but is shaped by the high variability within beginner and intermediate groups. These findings support the idea that technical vocabulary knowledge improves with proficiency level and vocabulary size, as advanced learners outperformed both intermediate and beginner learners. This is consistent with previous studies indicating that higher language proficiency correlates with a stronger grasp of specialised vocabulary, often due to increased exposure to technical content in both academic and professional settings. However, the significant variation within beginner and intermediate groups highlights the importance of targeted vocabulary instruction in engineering fields to bridge the gap between lower and higher proficiency groups, helping beginners and intermediate learners build foundational knowledge that supports their academic and professional success.

6.4.4 Engineering Teachers' Evaluation of Pedagogic Usefulness of ETV List

This section presents the results of engineering teachers' perceptions regarding the pedagogical usefulness of technical vocabulary items. Integrating corpus data with teacher evaluations of the pedagogic usefulness of words is an innovative approach adopted to enhance the practical relevance of word list studies (Dang et al., 2022a). Teacher feedback is crucial as it reflects the contextual and instructional realities within language classrooms, providing insight into how effectively corpus-based word lists can be applied to L2 teaching environments. The table below displays bootstrapped descriptive statistics, showing the proportion of teachers who rated the usefulness of words, categorized by word frequency levels.

Table 6-11: Teacher Evaluation of the Pedagogic Usefulness of Engineering TV list

		BCa 95% CI of mean		
Variables	**Mean**	**Lower**	**Upper**	**SD**
Proportion of teachers rating the words				
High-frequency (70 words)	236.61	195.60	279.48	180.97
Mid-frequency (70 words)	197.80	144.04	259.52	263.28
Low-frequency (70 words)	96.77	72.85	125.17	108.83
All target words 210	177.06	151.82	206.07	202.73

Abbreviations: BCa, bias-corrected and accelerated; CI, confidence interval

Table 6.11 presents data on teacher ratings of the usefulness of ETV, based on three frequency levels (high, mid, and low). The table uses bootstrapped descriptive statistics to analyse these perceptions across different frequency levels. Bootstrapped descriptive statistics were used to calculate means, standard deviations, and confidence intervals for the data collected through teacher surveys on the pedagogic usefulness of the engineering vocabulary list. A total of 210 words were carefully selected from 930 technical vocabulary items, with 70 words from each frequency band to ensure equal representation. For high-frequency words (N=70), teachers rated them with a mean score of 236.61, and a 95% CI ranging from 195.60 to 279.48. This suggests that teachers generally perceived technical vocabulary at the high-frequency level as highly useful, which aligns with expectations due to their frequent occurrence in language use. Examples of such words include ‘*safety*’, ‘*technology*’, ‘*structure*’, ‘*components*’, ‘*machine*’, ‘*capacity*’, and ‘*application*’.

Mid-frequency words (N=70) received a lower mean rating of 197.80, with a 95% CI from 144.04 to 259.52. This indicates that teachers still consider technical vocabulary at this level quite useful, though less so than those found in the high-frequency level. The wider confidence interval suggested more variability in teachers’ perceptions of mid-frequency words. Words under this category include *ceramic, turbine, torque, electromagnetic, armchair, vacuum,* and *condenser*.

In contrast, teachers rated the 70 technical vocabulary items in the low-frequency range considerably lower, with a mean score of 96.77 and a 95% CI from 72.85 to 125.17, compared to their counterparts in the high- and mid-frequency bands. This lower rating implies that teachers perceived low-frequency words as less useful, which is logical given their less common occurrence in everyday language. When considering all target words together (N=210), the overall mean rating was 177.06, with a 95% CI from 151.82 to 206.07. This average falls between the ratings for mid- and high-frequency words, reflecting the combined perceptions across all frequency levels. Notably, the SD for each category reveal interesting patterns. High-frequency words had an SD of 180.97, mid-frequency words 263.28, and low-frequency words 108.83. The substantially higher SD for mid-frequency words suggests greater disagreement among teachers about the usefulness of these words compared to high and low-frequency words. Some teachers may view mid-frequency words as important for their students' immediate learning, while others may consider them less essential. The perceived relevance of these words can vary depending on the specific engineering discipline or course content. For instance, terms like 'design' or 'optimization' may be central in disciplines such as mechanical or industrial engineering but less critical in fields like civil or environmental engineering (Coxhead, 2000). This highlights how the context and scope of a particular course can influence teachers' views on the importance of certain vocabulary.

Teachers' professional backgrounds may shape their perceptions of which vocabulary is most useful. According to Laufer (1997), teachers with extensive industry experience may prioritize practical, field-specific terms that students are likely to encounter in real-world engineering settings. In contrast, teachers with more academic experience may emphasize more specialised or technical terms crucial for academic writing or research. This variation is supported by Kang (2014), who found that different teaching methodologies and professional experiences can lead to divergent views on vocabulary importance. Teachers focusing on industry applications might prioritize functional vocabulary, whereas those emphasizing theoretical or academic instruction might focus on discipline-specific technical terms. Therefore, the perceived usefulness of mid-frequency vocabulary can vary significantly based on the teacher's professional background, the engineering discipline in question, and the context of the course, as discussed in the studies by Coxhead (2000) and Laufer (1997).

The findings reveal distinct patterns in how teachers value words for instructional purposes across frequency levels, indicating a stronger preference for high-frequency words. All target words were technical vocabulary identified in Study 2. An interesting finding from teachers' ratings of ETV usefulness is how they prioritize high-frequency technical vocabulary and consider low-frequency technical vocabulary less pedagogically essential. This is likely due to these words' prevalence and foundational role in technical communication. This is consistent with findings by Dang, et al., (2022a) that teachers prioritize vocabulary based on perceived instructional usefulness and frequency of real-world use. The variation in teachers' patterns of perceived word usefulness across frequency levels reflects instructional flexibility and the need to tailor vocabulary to specific learner levels. This aligns with the common assumption that words perceived as being useful for L2 learners may vary greatly between teaching contexts and vocabulary levels (Dang, et. Al, 2022a).

Table 6.12 below shows the top 20 most useful technical vocabulary items across three frequency levels (high, mid, and low) as rated by the teachers. It provides insights into which terms teachers find most relevant across these frequency categories, their perceptions of vocabulary usefulness, and their potential to support engineering education.

Table 6-12: Top 20 Most Useful Words at Each Frequency Level

SN	High-frequency words			Mid-frequency words			Low-frequency words		
	TV	Mean	SD	TV	Mean	SD	TV	Mean	SD
1	Heating	4.6	0.68	Generators	4.7	0.47	Setup	4.3	0.47
2	Machine	4.6	0.5	Solar	4.4	0.68	Exponential	4.1	0.96
3	Safety	4.6	0.68	Voltage	4.4	1.23	Workflow	4.1	0.96
4	Conductivity	4.4	1.23	Turbine	4.3	0.8	Inverter	4	0.79
5	Capacity	4.3	1.03	Fluid	4.3	0.47	Orthogonal	4	1.12
6	Estimation	4.2	0.61	Amplitude	4.3	0.8	Modulus	3.9	1.16
7	Loads	4.2	0.61	Valves	4.3	0.8	Neutron	3.9	0.96
8	Technology	4.2	1.19	Velocities	4.2	0.89	Impedance	3.8	1.28
9	Components	4.2	1.1	Axis	4.2	1	Desalination	3.8	1.19
10	Experiment	4.2	1	Composites	4.2	1	Seawater	3.8	1.19
11	Controller	4.1	1.16	Hardware	4.2	1	Throughput	3.8	1.36
12	Density	4.1	1.16	Vacuum	4.1	0.71	Axial	3.7	1.21
13	Waves	4.1	1.16	Entities	4.1	0.85	Wimax	3.7	1.21
14	Boundary	4.1	1.25	Friction	4.1	0.85	Bandwidth	3.7	1.3
15	Dimensional	4.1	0.96	Condenser	4	0.91	Hydropower	3.6	1.04
16	Filters	4	0.91	Rotational	4	0.79	Analog	3.6	1.31
17	Fibers	4	0.91	Compression	4	1.29	Linearization	3.6	1.31
18	Storage	4	1.12	Magnitude	3.9	0.55	Ergonomic	3.5	1.53
19	Structure	4	1.37	Electron	3.9	1.16	Hydroelectric	3.4	1.46
20	Quantity	4	1.37	Threshold	3.9	0.85	Neutrons	3.4	1.14

As shown in Table 6.12, the high-frequency words have consistently received high ratings, with average mean scores ranging from 4.2 to 4.6 for most of the words. This indicates that teachers consider them very useful in the engineering field. The standard deviations (SD) range from 0.5 to 1.37, showing moderate consistency in teachers' ratings. Words like '*heating*' and '*machine*' received high ratings with low standard deviation, indicating broad consensus on their usefulness. High-frequency words like '*safety*', '*technology*', and '*components*' are often considered foundational in engineering education due to their broad applicability across various disciplines and contexts. These terms are essential not only for understanding technical material but also for ensuring students' engagement in practical, everyday situations they will encounter in the professional world. For example, '*safety*' is critical across all engineering disciplines as it directly influences the design, construction, and operation of projects. Whether it pertains to electrical engineers considering safety protocols for machinery or civil engineers ensuring structural safety in building projects, this term holds universal significance. Similarly, words like '*technology*' and '*components*' transcend individual fields, as technology is at the heart of almost every engineering task, and components are fundamental in understanding how systems are constructed and work together.

The consensus among teachers regarding the importance of these high-frequency words suggests a shared understanding of the essential vocabulary that all engineering students must grasp early on. These words are likely seen as enabling students to both comprehend complex academic materials and participate meaningfully in professional engineering practices. For example, when students encounter terms like '*technology*' or '*components*' in various engineering contexts, they will be better equipped to understand technical specifications, equipment descriptions, and broader discussions around innovations and applications in engineering fields.

In the broader context of L2 vocabulary learning and teaching, this finding highlights the significance of prioritizing high-frequency terms in curricula. Research by Nation (2001) has shown that L2 learners need to acquire a substantial amount of high-frequency vocabulary to reach a functional level of language proficiency, particularly in specialised fields like engineering. High-frequency words often serve as building blocks for more specialised or complex terminology, allowing students to navigate texts more effectively. Thus, by ensuring that students develop a strong grasp of these foundational terms, instructors provide a solid base on which further language learning can be built.

Moreover, the broad agreement on the importance of these words in engineering education may reflect the universal nature of certain core concepts across disciplines. Teachers are likely to recognize that students' mastery of these terms will enable them to communicate effectively both within and outside academic settings, contributing to their professional development. As Schmitt (2008) emphasizes, vocabulary knowledge is crucial for learners to engage in both productive and receptive language tasks, with high-frequency words being particularly important for comprehension and fluency.

The slightly lower average ratings by teachers for mid-frequency words compared to high-frequency ones suggest that while these terms are useful, they are often more specialised, and their relevance may depend on specific contexts within engineering. Mid-frequency vocabulary typically bridges the gap between fundamental, everyday terms and the more highly specialised, technical terminology that advanced learners will encounter. As such, these words are important for students to understand as they begin to engage with more complex subject matter in their studies. For example, terms like '*generators*' and '*voltage'* are crucial for certain branches of engineering, particularly electrical engineering. '*generators*' may be particularly relevant for students studying power systems, energy generation, or machinery, whereas '*voltag*e' is a key term in the study of circuits and electrical theory. The mean score for '*generators*' (4.7, SD 0.47) suggests that teachers widely agree on its usefulness, likely because it is central to the study of electrical systems and is commonly used in both academic and practical contexts. On the other hand, '*voltage*' has a slightly lower mean score (4.4, SD 1.23), with a greater standard deviation, indicating some variation in how teachers perceive its utility. This variability may arise because while '*voltage*' is foundational in electrical engineering, its perceived importance might differ based on the specific course or study focus. For example, in a general introductory engineering course, '*voltage*' might be seen as essential, but in more specialised courses focused on mechanical or civil engineering, it may not hold the same immediate relevance. The standard deviations for these words (ranging from 0.47 to 1.23) further underscore this variability in teacher perceptions. This suggests that while there is general agreement on the usefulness of mid-frequency words, teachers' backgrounds, professional experiences, and teaching contexts may lead to differing views on the exact relevance of specific terms. Some teachers may prioritize terms that are central to their specific discipline, while others may view certain terms as more peripheral, depending on how frequently they appear in their teaching materials or real-world practice. This aligns with research by Schmitt (2000), who suggests that vocabulary relevance can vary across disciplines and even within subfields of a discipline, influencing how teachers evaluate the importance of specific words.

This finding has important implications for L2 vocabulary learning and teaching. It highlights the need for curriculum designers to consider not just the frequency of words but also the disciplinary context in which those words are used. Mid-frequency vocabulary, though not as universally applicable as high-frequency terms, plays an important role in deepening students' understanding of specialised topics. Teachers may need to strike a balance between teaching high-frequency words, which serve as foundational knowledge, and mid-frequency terms, which allow students to engage more deeply with specific subject areas.

In L2 vocabulary teaching, these findings suggest that a one-size-fits-all approach may not be the most effective. Teachers should consider the varying needs of their students, based on their academic focus and level of proficiency. For example, in a course focused on electrical engineering, teaching *'voltage'* and '*generators'* may take priority, whereas in a civil engineering context, words like '*load-bearing'* or 'stress' might be more pertinent. Therefore, mid-frequency words should be strategically incorporated into lesson plans, considering both their academic relevance and the likelihood that they will be encountered in real-world professional contexts.

Moreover, teachers should be aware that the perceived usefulness of mid-frequency words might change over time, as students' progress in their studies and gain more specialised knowledge. As learners advance, their vocabulary needs will shift, and what was once a specialised term may become a basic building block in their professional lexicon. This progression mirrors the model of vocabulary acquisition proposed by Nation (2001), where learners first acquire high-frequency words and then move on to mid- and low-frequency words as they gain expertise in a particular field. Thus, mid-frequency words play a crucial role in bridging the gap between general knowledge and highly technical language, making them essential for L2 learners aiming to become proficient in a specialised field such as engineering.

Finally, low-frequency words tend to have the lowest usefulness ratings, with mean scores averaging from 3.4 to 4.3. Teachers view these as less essential overall, likely because they are either advanced or less commonly encountered in general engineering contexts. The standard deviations for low-frequency words are higher (from 0.47 to 1.53), suggesting less agreement among teachers regarding these terms' usefulness. Technical words such as '*ergonomic'* (mean 3.5, SD 1.53) *and 'hydroelectric'* (mean 3.4, SD 1.46) show higher variability, indicating differing opinions on their relevance. Technical vocabulary at this level such as '*exponential', 'workflow', and 'neutron'* reflect specialised knowledge, showing that teachers consider these valuable primarily for specific fields within engineering rather than as foundational vocabulary.

6.4.5 The Relationship Between Learners' Vocabulary Knowledge and Teacher Perceptions Usefulness of ETV

To address Research Question 8, this section presents a correlational analysis examining the relationship between engineering teacher perceptions of the usefulness of engineering technical vocabulary and the receptive knowledge of engineering technical vocabulary among Saudi undergraduate engineering students. The results are presented in Table 6.13 below.

Table 6-13: Correlation Between Learner Vocabulary Knowledge and Teacher Perception of Word Usefulness

Teachers	**Groups of learners**			
	Low	Mid	High	All
N = 20	N = 19	N = 33	N = 26	N = 78
	0.72**	0.76**	0.813**	0.856**

** Correlation is significant at the 0.01 level (2-tailed).

Table 6.13 presents the correlation coefficients between learners' vocabulary knowledge (measured by the ETV test) and teacher perceptions of word usefulness across proficiency groups. These correlations indicate the extent to which teachers' perceptions reflect learners' actual vocabulary knowledge at different stages of development.

For low-proficiency learners (N=19), the correlation of 0.72 indicates a strong positive relationship, suggesting that teachers' judgments are reasonably well-aligned with the vocabulary knowledge of beginners. As proficiency increases, the correlations grow stronger: 0.76 for mid-proficiency learners (N=33) and 0.813 for high-proficiency learners (N=26). The strongest correlation is observed for the entire sample (N=78), at 0.856. These findings indicate that teachers' perceptions of word usefulness correspond more closely with learners' actual vocabulary knowledge as proficiency advances. In other words, teachers appear to have a more accurate intuitive understanding of the lexical knowledge of more advanced learners.

All correlations are statistically significant ($p < 0.01$), confirming the robustness of these relationships. This suggests that teacher judgments of word usefulness are reliably associated with learners' vocabulary levels. This does not mean that teachers are explicitly estimating which words learners know; rather, their perceptions of usefulness align with the vocabulary breadth learners demonstrate, particularly at higher proficiency levels.

This pattern aligns with previous research, such as Dang et al. (2022b), who also reported a significant relationship (e.g., $r = 0.67$) between teacher perceptions and learner knowledge. This reinforces that teachers' perceptions, especially when they are familiar with the learner context, tend to correspond closely with actual lexical development.

In sum, the results demonstrate a clear and strengthening relationship between teacher-judged word usefulness and learner vocabulary knowledge across proficiency levels. The statistically significant correlations, which increase from low to high proficiency groups, indicate that teachers' perceptions are not arbitrary but are closely attuned to the lexical realities of their learners. This provides empirical support for the value of teacher expertise, suggesting that their intuitive judgments reliably reflect the vocabulary that learners are likely to know at various stages of their development.

6.5 Summary

Study 3 aimed to connect corpus-based word lists with actual ESP vocabulary learning contexts, addressing two research questions: the extent of Saudi engineering students' receptive knowledge of technical vocabulary (RQ4) and the vocabulary items teachers perceive as most pedagogically useful (RQ5). This study bridged the gap by incorporating end-user perspectives, specifically assessing learners' recognition of technical vocabulary alongside teacher evaluations of these words' instructional value.

The study confirmed that corpus data alone cannot fully capture the instructional relevance of vocabulary in specific fields. Teacher evaluations underscored the pedagogical and contextual significance of technical terms in engineering. Yes/No test results revealed that students have strong receptive knowledge of high-frequency technical vocabulary, with higher scores and lower variability at this level, indicating solid mastery. However, recognition declined significantly at mid- and low-frequency levels, where guessing increased, showing limited exposure to less common vocabulary. This trend suggests that high-frequency terms should be mastered before introducing mid- and lower-frequency items to enhance learning effectiveness.

Teacher evaluations aligned with this approach, prioritizing high-frequency vocabulary for its wider relevance in technical communication and considering lower-frequency words less critical. Teachers' feedback provides valuable insight into classroom applicability, illustrating how corpus-based word lists can be adapted for language instruction. This study demonstrates the importance of aligning technical vocabulary with practical, pedagogically relevant criteria to optimize ESP teaching and learning outcomes in engineering contexts.

Chapter 7 : Discussion and Conclusion

7.1 Introduction

Chapters 4, 5, and 6 presented detailed accounts of three interconnected studies, each addressing the research questions and reporting key findings. This chapter offers a comprehensive discussion of these findings, structured around three main themes: (1) an overview of engineering technical vocabulary (ETV) in academic contexts; (2) the development of discipline-specific word lists and multi-word units; and (3) students' knowledge of, and teachers' perceptions about, ETV. While the first theme represents a central focus of the thesis, it also informs and supports the second, as both contribute to understanding how technical vocabulary is identified and categorized within the engineering domain. Together, these two themes establish the foundation for the third theme, which situates ETV within the wider context of vocabulary learning and teaching, providing insights relevant to both engineering education and ESP. Each theme is explored in detail in the sections that follow.

7.2 The Lexical Profile Analysis of Engineering Masters' Dissertations Corpus (EMDC*)*

As discussed in Chapter 4, the first part of Study 1, applied LFP and examined the coverage of Schmitt and Schmitt's (2014) high-, mid-, and low-frequency vocabulary in the EMDC. This methodological framework was used in previous lexical profile studies in ESP such as Dang (2020), Dang et al. (2022), and Lu (2018). The analysis was conducted using AntWordProfiler, a robust tool designed to calculate the lexical profile within texts. This analysis provided a detailed breakdown of vocabulary distribution in the EMDC across Nation's (2012) BNC/COCA word lists, plus supplementary word lists. The key findings addressing Research Question 1 are summarized below.

To begin with, the lexical profile of the EMDC reflected the patterns observed in previous profiling research, such as those conducted by Dang (2020) and Dang et al., (2022a). The lexical profiling of the EMDC revealed that the cumulative coverage of the BNC/COCA base word lists (levels 1 to 25) accounted for only 93.3% of the tokens, falling short of the 95% minimum threshold for acceptable reading comprehension (Laufer, 1989). This part further explored the importance of each frequency band in the context of engineering masters' dissertations. The results obtained through the LFP analysis in this thesis clearly demonstrate the vocabulary coverage of engineering masters' dissertations across Nation's (2012) BNC/COCA frequency bands. The results revealed word distribution patterns similar to those reported in previous studies, such as those by Benson and Coxhead (2022), Lu (2018), and Lu and Coxhead (2020). The first 1,000 base-level word list contained the highest percentage of tokens in the EMDC (59.26%), with a noticeable decrease in coverage in the second and third 1,000-word lists (14.82% and 11.25%, respectively), continuing through to the 25th 1,000 base word lists. These findings indicate the inherent skewness of vocabulary distribution across levels (Durrant, 2016). Moreover, Nation and Waring (1997) advocate that general English learners should first focus on the first 3,000 high-frequency words, as they serve as a foundation for basic text comprehension.

Understanding the role of high-, mid-, and low-frequency vocabulary is essential for end users of the texts seeking to enhance the quality and readability of these dissertations. The study revealed that high-frequency word families, encompassing the first 3,000-word families, made up 85.33% of the running words in the EMDC. This emphasized the important role of high-frequency vocabulary, which provided the majority of lexical coverage, enabling students to comprehend texts across academic disciplines. This aligned with Dang's (2020) work, which highlights the critical importance of high-frequency words in facilitating academic discourse comprehension.

Similarly, mid-frequency vocabulary, comprising 4,000–8,000-word families, accounted for 6.6% of the EMDC. Although less common than high-frequency words, these terms are pivotal for understanding technical discussions in specialised domains. Dang (2020) and Lu (2018) emphasize the bridging function of mid-frequency vocabulary, linking general academic language to highly specialised terminology. For engineering students, familiarity with mid-frequency vocabulary is crucial for comprehension of core engineering concepts and technical discourse effectively. Furthermore, low-frequency vocabulary, representing 9,000–25,000-word families, contributed a modest 1.76% of the total coverage in the EMDC as in previous vocabulary studies in ESP (e.g., Lu, 2018; Benson, 2020). As noted by Schmitt and Schmitt (2014) and corroborated by Dang, et al., (2022a), the occurrence of low-frequency vocabulary in specialised texts is highly contextual. In the EMDC, this reflects the specialised nature of engineering discourse, where such vocabulary is essential for precise technical expression but not for general comprehension.

The result obtained also revealed that supplementary word lists contributed 6.16% to the total coverage, highlighting their critical role in bridging the gap between general vocabulary and more specialised terminology. These supplementary lists are especially important because they account for proper nouns, abbreviations, and terms that, while not frequent in general language use, are essential for understanding specific academic or professional texts (Schmitt & Schmitt, 2014).

Another interesting result from the frequency analysis was the development of possible technical words in the EMDC, following the methodological framework adopted by previous vocabulary studies in ESP, such as that of Benson (2020), Lu (2018), Lu and Coxhead (2020) and Benson and Coxhead, (2022). Most of the words in the Engineering Base Word list that were not included in the BNC/COCA frequency base lists were loan words, such as *microstrip* (derived from Greek *mikros*, meaning '*small*', and Latin *stripa*, meaning '*strip*'), and *rotable* (from Latin *rota*, meaning '*wheel*'). These words are known as fully technical words (Lu, 2018) or domain-specific vocabulary (Fraser, 2013) and often originate from Latin or Greek. The substantial coverage of such words may pose difficulties in comprehending the EMDC for EFL learners. This result emphasized the need for dedicated technical vocabulary lists tailored to the field of engineering, which required a more targeted approach to ensure adequate comprehension of engineering texts. These findings align with research on specialised vocabulary acquisition, which suggests that learners need access to both high-frequency general vocabulary and discipline-specific terms to fully engage with professional texts (Lu, 2018; Benson, 2020). As highlighted by Benson and Coxhead (2022), the ability to understand and use technical vocabulary is critical for students to succeed in their field. Therefore, the inclusion of engineering-specific word lists alongside general vocabulary lists is vital for preparing students to navigate and comprehend the technical language of their discipline effectively.

7.3 Vocabulary load of the EMDC

This section presents key findings in relation to the vocabulary load analysis, which aimed to address research question 2 by providing the number of word families a learner needs to know to comprehend the EMDC. Two prominent vocabulary thresholds were established, as reported in the literature: a minimum of 95% and an optimal of 98% for vocabulary load coverage, as identified through corpus-based research on vocabulary size (Durovic, 2021; Coxhead and Demecheleer, 2018).

As discussed in Chapter 4, vocabulary load is the percentage of words in a text that learners are likely to know (Nation, 2013), as determined through the concept of 'coverage'. Vocabulary load is calculated by progressively summing the lexical coverage provided by each

BNC/COCA base word list until reaching established thresholds of 95% (minimum) coverage or 98% (optimal) coverage. These thresholds have been widely recognized as critical for adequate text comprehension (Laufer, 1989; Laufer & Ravenhorst-Kalovski, 2010; Hu & Nation, 2000; Van Zeeland & Schmitt, 2013). This helps to determine whether a text is at the right level for learning (Webb & Nation, 2008). Understanding coverage is crucial for determining how much vocabulary learners should know to achieve adequate reading comprehension of texts (Schmitt, Jiang, & Grabe, 2011). The key findings addressing research question 2 are summarized below.

The result of vocabulary load analysis revealed that vocabulary sizes of 5,000 and 7,000 words were required to reach the minimal and optimal threshold levels of 95% coverage and 98% coverage in the EMDC, respectively. This means that the comprehension of engineering masters' dissertations requires knowledge of the first five 5,000- and 7,000-word families together with knowledge of supplementary lists and additional engineering word lists for learners to attain minimal and optimal reading comprehension. Interestingly, the vocabulary load results revealed that the EMDC is lexically demanding to comprehend. Henceforth, the vocabulary load analysis highlighted the lexical demand of the EMDC, drawing comparisons with findings from studies on vocabulary in other specialised genres and disciplines.

These findings are consistent with previous research on vocabulary requirements in various academic and technical domains. For example, Nation (2006) found that 8,000-word families plus proper nouns were necessary to achieve 98% coverage in university textbooks, novels, and newspapers. Similarly, Coxhead and Boutorwick (2018) reported that Grade 8 Math textbooks required 8,000-word families, along with proper nouns, compounds, and abbreviations, to meet the 98% comprehension threshold. Coxhead et al. (2016) identified that carpentry texts required 8,000-word families plus discipline-specific vocabulary, while Hsu

(2014) found engineering textbooks required around 5,000-word families for university-level comprehension. In L1 and L2 graduation theses, Stamatović et al., (2020) noted that 95% coverage was achieved at 4,000-word families with supplementary lists, and 98% coverage required 9,000-word families. Hsu (2013) found that medical textbooks demanded a vocabulary load of 14,000-word families plus technical terminology for optimal comprehension. Other studies, such as Sun and Dang (2020), highlighted variations in lexical demands, with Chinese high school textbooks requiring 3,000-word families for 95% coverage and 9,000 for 98%. Similarly, Hsu (2018) reported that Traditional Chinese Medicine texts required 7,000-word families plus proper nouns for 95% coverage and 10,000-word families plus proper nouns for 98%.

In summary, the vocabulary load of the EMDC, as supported by prior studies, indicates that dissertations, as specialised written academic text genres, typically require a more extensive lexical knowledge than general academic texts. Therefore, EFL students may require rigorous vocabulary training to fully comprehend engineering-specific texts. Since the present study is mainly concerned with technical vocabulary, the next section addresses Study 2, which examines engineering technical vocabulary in the EMDC.

7.4 the Development of Single-word unit Engineering Technical Vocabulary Items

This section presents the key findings on the development of single-word engineering technical vocabulary items in the EMDC, addressing research question 3. As discussed in Chapter 5, this study adopted both corpus-based and semantic-based approaches to identify technical vocabulary in the EMDC, drawing on methodologies similar to those employed in recent research on technical vocabulary across various specialised disciplines and genres, such as Benson and Coxhead (2022), Hsu (2018), Coxhead et al. (2016), and Liu and Lei (2020). The key findings addressing Research Question 3 are summarized below.

One important finding is the development of 930 single-word engineering technical vocabulary items for pedagogical purposes. This was achieved using a combination of keyword analysis, frequency thresholds, dictionary consultation, and semantic rating scales. Initially, 2,470 keywords were identified, which were refined through frequency criteria and semantic analysis, resulting in 930 technical words. These included commonly used engineering terms such as '*figure*', '*system*', '*model*', and '*data*'. The final list was confirmed by specialists in the engineering field.

Another interesting finding is that the study examined the coverage of these technical vocabulary items, which accounted for 19.92% of the total tokens in the EMDC. While this is substantial, it is lower than the 33.5% coverage reported in Lu's (2018) study on Traditional Chinese Medicine. This difference may be attributed to variations in corpus size and the nature of technical vocabulary across different disciplines. Although different definitions and methods have been used by scholars to identify technical vocabulary, their research has consistently shown that technical terms make up a significant portion of specialised texts.

The finding of this study also describes how many of the engineering technical vocabulary belong to the high-, mid-, and low-frequency vocabulary bands, respectively. The findings revealed that 570 out of the 930 identified technical words were found within Schmitt and Schmitt's (2014) 3,000 high-frequency vocabulary bands. This indicates a substantial number of ETV were from general high-frequency vocabulary, which highlight the importance of general high-frequency vocabulary in engineering domain. Similarly, in the rugby domain, Benson (2020) found that substantial number of technical words of both spoken and written technical single-word lists in rugby corpora were from 3,000 high-frequency vocabulary band. Across specialised disciplines, research consistently shows that high-frequency generic vocabulary often takes on technical meanings. Tiersma (1999) initially reported this phenomenon in legal discourse, and further research has confirmed comparable trends in engineering (Watson Todd, 2017), pharmacology (Fraser, 2009), and medicine (Hsu, 2013; Quero & Coxhead, 2018). Together, these results imply that common lexical elements are frequently appropriated and repurposed for technical conceptualization in domain-specific communication.

Henceforth, this thesis aligns with the broader understanding that general high-frequency vocabulary can also serve technical functions across specialised domains, as observed in fields such as medicine (Lu, 2018; Quero & Coxhead, 2018), pharmacology (Fraser, 2005, 2009), and applied linguistics (Fraser, 2005). This also aligns with the conceptualization of technical vocabulary in the present thesis, where words with specialised meanings may also appear in the GSL or AWL (Tongpoon-Patanasorn, 2018). In this thesis, the terms 'specialised vocabulary' and 'technical vocabulary' are used interchangeably to reflect this nuanced understanding of technical language. Technical vocabulary is not confined to any specific frequency range; it can occur in high-, mid-, or low-frequency word lists (Benson, 2020; Benson & Coxhead, 2022; Chung & Nation, 2003; Nation, 2013, 2016). For example, everyday words like '*flow*' may have specialised meanings in fields such as plumbing (Coxhead & Demecheleer, 2018).

This emphasizes the importance of Saudi EFL engineering students acquiring discipline-specific vocabulary in order to fully comprehend texts related to their field. Hence, engineering EFL courses should integrate specialised technical word lists, such as those revealed in this study, to assist students in understanding and learning domain-specific vocabulary. Teachers might also draw attention to polysemous words that have both general and specialised engineering meanings. Overall, the study provided a comprehensive approach to identifying technical vocabulary in the EMDC, highlighting the significant role of both high-frequency and specialised vocabulary. This approach followed the criteria used by Benson (2020) and Lu (2018), which involved adopting Nation's (2012) base word levels 1–25 and conducting qualitative analysis using a semantic rating scale. This contrasts with the methodological approaches used in some prior studies, such as Wang et al. (2008), who identified technical vocabulary based on frequency, range, and specialised occurrence, excluding words from the GSL but including terms from the AWL. Similarly, Hsu (2013) established technical vocabulary using criteria such as specialised occurrence, range, and word family frequency, excluding terms from the BNC 3000 list while retaining AWL items. Thus, the framework used in the present study contributes an alternative perspective to technical vocabulary identification by leveraging Nation's (2012) comprehensive lexical classification system.

The present study also adopted keyword analysis to help identify potential technical vocabulary, based on the assumption that technical terms occur more frequently in specialised contexts than in general English (Chung, 2003; Scott, 2006). Moreover, the study did not apply range and dispersion as selection criteria, following a similar trend in several recent corpus studies focused on developing discipline-specific word lists (e.g., Benson & Coxhead, 2022; Durovic, 2021; Lu and Coxhead, 2020).

7.5 Multiword Units Engineering Technical Vocabulary Items

Building on the development of multiword units ETV, this section presents the key findings on multiword engineering technical vocabulary items within the EMDC, addressing research question 4. As discussed in part 2 of Chapter 5, the second aim of Study 2 is to develop MWUs from the EMDC. For this purpose, corpus-based and semantic-based approaches were applied to develop MWUs. The key findings addressing Research Question 4 are summarized below.

Firstly, the findings showed that 856 MWUs were initially identified from the EMDC and later refined to 543 condensed MWU technical vocabulary items based on Wood and Appel's (2014) root structures, as used in other studies on ESP (Benson, & Coxhead, 2022; Benson, 2020; Coxhead et al., 2017; Lu 2018). These MWUs ranged from two- to five-word combinations, with key phrases like '*energy source'* and '*boundary condition'*. A root structure principle was applied to manage overlapping multiword units, resulting in 543 stand-alone structures and additional variable-slot structures.

Another important finding is the dominance of nominal structures: the most common pattern in the MWUs of the EMDC was nominal structures, with 539 nominal phrases identified (e.g., '*energy source'*, '*membrane channel'*). Only a few multiword units were verb or prepositional phrases. This dominance of nominal structures aligns with findings from Lu's (2018) study and reflects the information-dense nature of engineering dissertations as a specialised academic genre.

These phrases often serve to name complex phenomena, entities, or processes (Ward, 2007). Similarly, Salager (1983) observed frequent use of compound noun phrases in medical English. Ward (2007) developed a multiword item list called the Technical Collocations List, which found 78 noun phrase collocations (4 with four words, 19 three words, and 55 two words). In contrast to the present study, Ward (2007) employed only a frequency cut-off point, focusing on noun phrases that appeared three times or more in the corpus to be selected for inclusion in the list.

Furthermore, the findings revealed that the majority (69.2**%)** of MWUs in the EMDC (540 items) were two-word patterns, highlighting the significance of these shorter combinations in engineering discourse. This aligns with the recommendations of researchers such as Benson and Coxhead (2022) and Lu (2018), who emphasize the pivotal role of two-word units in academic and technical language. Similarly, Benson (2020) observed that most root structures consist of two words, with 154 of the 174 stand-alone MWUs in the word list being two-word units. As detailed in Chapter 5, a targeted list of 543 condensed MWU technical vocabulary items was created by methodically reducing the original 856 MWU items. A methodological strategy frequently used in ESP research (Benson, 2020; Benson & Coxhead, 2022; Coxhead et al., 2017; Lu, 2018) was applied to achieve this reduction: Wood and Appel's (2014) root structure framework. While preserving the technical vocabulary list's disciplinary relevance, the simplified inventory enhances its educational usefulness.

In consistent with prior studies, a considerable number of two-, three, and four-word MWUs are incorporated within three-, four-, and five-word forms. To address issues of overlap, these were condensed into simplified forms, as suggested by previous research (Benson, 2020; Benson & Coxhead, 2022; Coxhead, 2019; Byrd & Coxhead, 2010; Wood & Appel, 2014).

Overall, this thesis emphasizes the crucial role of MWUs in the engineering discipline, highlighting the prerequisite need for learners to acquire both single-word and multiword technical terms in the engineering discipline. While the combined lists of technical MWUs and single-word items present valuable learning targets for engineering students, effectively mastering such extensive vocabularies remains a significant challenge. The MWU engineering technical vocabulary list developed in this study is specifically tailored for pedagogical purposes at the university level and designed to support engineering students in achieving academic and professional success.

The engineering discipline-specific vocabulary lists created in this study, which include both single-word and multiword unit items, were intended for pedagogical use in university-level engineering education. These carefully selected lexical resources support students' academic success in engineering degree programmes and help prepare them for professional communication within the industry. Specifically designed for use in ESP classrooms, these lists provide teachers with focused instructional resources and offer students a step-by-step guide to learning key technical terms.

7.6 The Coverage of Engineering Technical Vocabulary Across the Five Main Sections of Masters' Dissertations

This section presents the key findings regarding to the coverage of engineering technical vocabulary across the five main sections (Introduction, Literature review, Methodology, Results and Discussion, and Conclusion) of engineering masters' dissertations. As detailed in part 3 of Chapter 5, the research question 5, concerns with examining the distribution of ETV across the five main sections of masters' dissertations. For this purpose, the EMDC was divided into five sub-corpora, each corresponding to one of the main sections: Introduction, Literature Review, Methodology, Results and Discussion, and Conclusion. Subsequently, a frequency profile analysis was conducted using AntWordProfiler to determine the coverage of the final list of single-word unit technical vocabulary items within each section of the EMDC. This analysis provided insights into the extent to which engineering technical vocabulary is represented across the different sections, offering a broader perspective on their use throughout the dissertations. The findings of this study highlight the varying coverage of engineering technical vocabulary across different sections of masters' dissertations in engineering, offering insights into how specialised language is employed within different stages of academic writing. The key findings addressing Research Question 5 are summarized below.

The introduction part of the EMDC contained 285 words from the 930 single-word unit technical vocabulary items, accounting for 30.6% of the word list. This suggests that, while technical vocabulary is present in the introduction, its usage is relatively limited compared to other sections, which may be due to the more general and contextual nature of the introduction.

The literature review part consisted of 441 words (47.4% of the 930-word list), indicating a substantial presence of technical vocabulary. This finding aligns with the typical function of the literature review, which involves discussing specialised topics, reviewing previous studies, and positioning the research within the existing body of knowledge. These aspects necessitate the use of domain-specific terms, as also observed by Thompson (2005), who found that intertextual references and disciplinary language are critical in positioning the researcher's work in PhD theses.

The methodology section contained 509 words, representing 54.73% of the technical vocabulary list. The substantial use of technical vocabulary in this section reflects the detailed description of procedures, instruments, and technical processes that characterize engineering research. This aligns with findings from similar studies, such as those by Thompson (2005), which demonstrated that the methodology section is typically rich in procedural and technical vocabulary that aids in clearly describing the research process.

The results and discussion section included 471 technical words, accounting for 50.65% of the list. This level of technical vocabulary usage is expected, as this section involves explaining and interpreting the findings using specialised terms that pertain to engineering concepts and measurements. The coverage here emphasizes the role of technical language in accurately discussing outcomes and drawing relevant conclusions based on data. Interestingly, the Conclusion section of the dissertations had the least coverage of technical vocabulary, with only 121 out of 930 words (13%), which could be attributed to the brief nature of many Conclusion sections. In the EMDC, the Conclusion section contained only 7,073 tokens, which may have limited opportunities to employ a wide range of technical terms. The Conclusion section typically summarizes key findings and discusses implications rather than delving into specific technical details, which would naturally result in reduced technical vocabulary usage.

In summary, the coverage of the engineering technical vocabulary varied across the macrostructures of masters' dissertations, reflecting the distinct functions and requirements of each section. These findings corroborate the observations of Thompson (2005), who highlighted those different sections of academic writing serve unique roles and therefore exhibit varying lexical characteristics. Understanding how technical vocabulary is distributed across different parts of a dissertation can provide valuable insights for ESP instruction, particularly in assisting non-native English-speaking students in effectively using domain-specific language throughout their academic writing.

One interesting finding of this thesis is the examination of the distribution of ETV across the overall discourse structure of dissertations, a new approach in genre analysis from the perspective of LFP, which remains underexplored in prior genre-based research on dissertations and theses (e.g., Dudley-Evans, 1999; Paltridge & Starfield, 2020; Sun & Crosthwaite, 2022; Thompson, 1999, 2005). The genre-based approach has played a key role in advancing research on disciplinary discourses, as advocated by corpus linguists (e.g., Hyland, 2000; Thompson, 2016). Hyland (2000), in particular, examines a range of text genres across disciplines, including soft and hard sciences and applied linguistics. This is important because texts are constructed to be interpreted within specific cultural contexts, and genre analysis can reveal the underlying norms of these academic cultures. Conventional rhetorical practices often reflect individual writers' perceptions of shared values (Hyland, 2010). Consequently, genre theorists emphasize the centrality of participants' interactions in language use, if an effective text will demonstrate the writer's awareness of its context and its audience (Hyland, 2005). In addition, the present thesis suggests that understanding the lexical distribution across the sub-sections of the dissertations is important, as it involves characterizing text variation based on organizational patterns, a key concern of genre analysis, as described by Biber and Conrad (2009). This implies that lexical analysis of texts from a genre perspective centres on the conventional frameworks used to construct complete academic texts. Consequently, the genre-based approach in this study emphasizes detailed and comprehensive text analysis (Egbert & Gray, 2019). Moreover, the present thesis views dissertations as relatively less interactive and multimodal research genres within academic discourse (Hyland & Shaw, 2016). Postgraduate research students, in particular, are expected to produce texts like research articles and conference papers before finalizing their theses (Hyland & Shaw, 2016).

7.7 Saudi undergraduate Students' Receptive Knowledge of Engineering Technical Vocabulary

This section presents the key findings on Saudi undergraduate engineering students' receptive knowledge of ETV items, addressing research question 6. As discussed in part 1 of Chapter 6, this thesis incorporates information from end-users both teachers and learners to complement the corpus-based generated ETV lists developed in the prior Study. This aligns with prominent word list researchers such as Dang and Webb (2016) and Dang, et al., (2022a), who used learners' information to supplement corpus data about general high-frequency word lists. As a key element of language learning, effective vocabulary learning is necessary for integration into a technical community (Nation, 2013). Learners must develop both receptive and productive knowledge of technical vocabulary in their respective fields of study. The study examines Saudi undergraduate engineering students' receptive knowledge of technical vocabulary in their field of study. The present study involved 78 final-year students from NBU, Saudi Arabia, majoring in Electrical, Mechanical, Civil, and Industrial Engineering. The key findings addressing Research Question 6 are summarized below.

First, UVLT (Webb, et al., 2017) was used to assess the participants' general vocabulary level knowledge. They were categorized into three groups: High (H), Mid (M), and Low (L) vocabulary proficiency. Results showed that the H group, comprising 19 participants, had a solid grasp of high-frequency vocabulary (1K–3K) but struggled with less frequent vocabulary (4K–5K). The M group, with 33 participants, had mastered the most frequent words (1K–2K) but lacked knowledge beyond the 3K level. The L group, consisting of 26 participants, had a limited vocabulary, lacking knowledge even of the most basic English words (1K–2K). These results highlight varying vocabulary proficiency levels and the need for differentiated instruction to address each group's needs. However, the study concludes that Saudi undergraduate engineering students tend to acquire and retain high-frequency words more easily than low-frequency words due to their greater exposure to these words in everyday language use. This aligns with lexical profiling studies in ESP, where learners exhibit stronger receptive knowledge of the most frequent word bands (e.g., the 2K band) but struggle with less common words (Dang, 2017; Webb & Chang, 2012). Nation (2001) attributes this to learners' natural prioritization of high-frequency words for foundational comprehension before advancing to specialised vocabulary.

The findings showed that participants' knowledge of technical engineering vocabulary was assessed using three Yes/No tests. Each test included 70 real technical words and 30 pseudowords, drawn from high-, mid-, and low-frequency bands of the 930 technical words identified in the study's corpus. The scoring followed established methods (Mochida & Harrington, 2006; Meara, 1989; Huibregtse et al., 2002), including hit rates, false alarm rates, and correction formulas. Key findings showed that participants were more accurate in identifying high- and mid-frequency technical words than low-frequency words. Variability in performance was greater for low-frequency words, indicating that students struggled more with less familiar terms. Advanced learners (N=26) had the highest receptive knowledge of technical words (62.81%), followed by intermediate learners (N=33, 53.33%), and beginner learners (N=19, 50.32%). These findings reflect the typical patterns of receptive vocabulary knowledge commonly found among EFL learners. For instance, Webb and chang (2012) found that only 26 advanced learners had mastered the 3K level, suggesting that even proficient students may struggle with lower-frequency vocabulary, as posited by Laufer and Nation (1999). These results emphasize the need for differentiated vocabulary instruction tailored to learners' specific proficiency levels (Nation, 1990). Finally, this study highlights the importance of aligning corpus-based word lists with real-world vocabulary learning needs to enhance students' comprehension of technical terms. As supported by prior studies such as Lui and Lei (2020) and Knoch (2014), receptive knowledge of technical vocabulary is essential not only for mastering subject content but also for developing effective professional communication skills within a specialised discipline.

7.8 Teachers' Perceived Pedagogical Usefulness of Engineering Technical Vocabulary

This section presents the key findings on engineering teachers' perception of pedagogical usefulness of engineering technical vocabulary, through their evaluation of ETV list, addressing research question 7. As discussed in part 2 of Chapter 6, this thesis incorporates information teachers' evaluation of pedagogical usefulness of ETV to complement the corpus-based generated ETV lists developed in the Study 2. The key findings addressing Research Question 7 are summarized below:

The first key finding regarding teachers' evaluation of 210 engineering technical vocabulary items, categorized according to Schmitt and Schmitt's (2014) frequency bands, shows clear patterns in teachers' perceptions of the pedagogical usefulness of ETV. High-frequency words (N = 70) received the strongest endorsement, with a mean rating of 236.61 (95% CI: 195.60–279.48), reflecting broad consensus about their essential role in instruction, as evidenced by a moderate standard deviation (180.97). This finding suggests that teachers generally perceive high-frequency words as highly useful, which aligns with expectations given their prevalence in both general and disciplinary language.

The findings on mid-frequency technical vocabulary reveal several intriguing trends in teachers' perceived usefulness of ETV. With a mean usefulness rating of 197.80, considerably higher than low-frequency terms but notably lower than high-frequency items, these results suggest that mid-frequency vocabulary occupies a large, yet challenging, middle ground in terms of instructional value. There is significantly less agreement among instructors regarding the usefulness of these items, as indicated by the 95% confidence interval of 144.04 to 259.52 (a span of 115.48 points, compared to only 83.88 for high-frequency words). This significant variety likely reflects numerous underlying elements in engineering education contexts. The inconsistent rating of mid-frequency words gains further significance when examining specific examples from this group. For instance, faculty in mechanical engineering fields may routinely give high marks to words such as '*turbine*' and '*torque*', which are essential, whereas faculty in electrical or chemical engineering fields may view these terms as less important. In contrast, '*electromagnetic*' may receive lower ratings from civil engineering faculty but be considered essential by those in electrical engineering. This disciplinary variation naturally influences perceived usefulness, as reflected in the broader confidence intervals. Similarly, words like '*ceramic*' and '*condenser*' highlight how application-specific utility impacts ratings: a software engineering instructor may view '*ceramic'* as peripheral, while a professor of materials science might consider it important. The case of '*vacuum*' is especially instructive because, although it is essential in some mechanical and physical applications, it may be viewed as less important in many other engineering contexts. Consequently, the standard deviation for mid-frequency words (263.28) was significantly higher than that of both high- and low-frequency items, reinforcing the variability in their perceived instructional value.

The above findings imply that, while mid-frequency technical vocabulary is widely acknowledged as significant, its educational value is highly context-dependent. Unlike high-frequency terms, which tend to transcend specialisations, or low-frequency words, which are consistently niche, mid-frequency vocabulary occupies a '*grey zone*' where its usefulness is dictated by specific disciplinary needs and instructional contexts. This has significant ramifications for curriculum development, indicating that rather than being viewed as generally applicable, mid-frequency word lists may need to be modified or added to in accordance with particular engineering subfields. The findings highlight the necessity of adaptable, discipline-aware methods of teaching technical vocabulary that take into account the variation in perceived utility across various instructional settings.

Another important finding reveals that teachers systematically perceived low-frequency words as having significantly lower instructional value. This is evidenced by the substantially lower mean rating for low-frequency words (96.77) compared to both high-frequency (236.61) and mid-frequency (197.80) terms. This finding aligns with well-established vocabulary acquisition principles; teachers naturally place less instructional emphasis on low-frequency terms, which are, by definition, less common in both general and specialised discourse. Despite representing a variety of engineering fields, teachers' surprisingly high level of agreement regarding low-frequency words' limited utility is indicated by their moderate standard deviation (108.83) and rather narrow confidence interval (72.85-125.17).

It is crucial to consider the overall mean rating of 177.06 (95% CI: 151.82-206.07) when analysing the composite scores for all 210 vocabulary items. Positioned between the mid- and high-frequency word ratings, this intermediate value illustrates how frequency bands collectively influence teachers' perceptions of pedagogical usefulness. The range of the confidence interval (54.25 points) indicates moderate variability in teachers' overall evaluations of the usefulness of ETV items. This variation is primarily due to the combined effects of the considerably more variable mid-frequency assessments and the more stable ratings at the frequency extremes (low and high). Particularly useful insights into teacher agreement levels can be found in the standard deviation patterns. Mid-frequency vocabulary elicited significantly greater disagreement among teachers, as evidenced by the high SD for mid-frequency words (263.28), which is more than twice that of low-frequency words (108.83) and over 45% higher than that for high-frequency terms (180.97). This disparity is theoretically sound, as mid-frequency technical terms are likely essential in some engineering disciplines but peripheral in others. Conversely, both low-frequency words (due to their general obscurity) and high-frequency words (due to their universal importance) produce more consistent ratings across various instructional contexts.

Overall, the above findings show distinct trends in engineering teachers' perceptions of the pedagogical usefulness of ETV across different frequency bands. In particular, teachers consistently regarded higher-frequency ETV items as more helpful, while lower-frequency words were rated much less favourably. Words with high lexical coverage tend to appear more frequently and in a wider range of contexts, making them more likely to be acquired earlier than less frequent words (Ellis, 2002). Consequently, these words are perceived as more pedagogically useful by teachers.

7.9 The Relationship Between Engineering Teachers' Perceived Usefulness of Technical Vocabulary and Students' Receptive Knowledge

This section presents the key findings on the relationship between Engineering teachers' evolution of pedagogical usefulness of ETV items and students' receptive knowledge, addressing research question 8. As detailed in Chapter 6, a bootstrapped Pearson product-moment correlation was conducted to examine the potential relationship between learner-receptive vocabulary knowledge and teacher perceptions of the pedagogical usefulness of the ETV list. This method mirrors the approach used in He and Godfroid's (2019) study, which explored the relationship between frequency and teacher perceptions of academic vocabulary usefulness. The interpretation of correlation values (r) followed Plonsky and Oswald's (2014) effect size benchmarks for L2 research, where r values around .25 indicate a small effect, .40 a medium effect, and .60 a large effect (p. 889). The key findings addressing Research Question 8 are summarized below.

The key findings indicate a significant, proficiency-dependent relationship between teachers' perceptions of vocabulary usefulness and students' receptive knowledge of engineering technical vocabulary items. A significant overall correlation of 0.856 across all competence levels shows that teachers generally have accurate intuitions about the words that students are likely to know. This alignment becomes even more noticeable at higher proficiency levels, with correlations increasing from 0.72 for low-proficiency learners to 0.813 for advanced students. According to this pattern, learners' vocabulary knowledge more closely resembles the lexical distribution that teachers naturally consult when assessing a word's utility as they gain more language proficiency.

Advanced students begin to comprehend words in distributions that more closely resemble teachers' perceptions about useful technical language when they have progressed beyond basic vocabulary, as supported by Nation (2022). On the other hand, beginners' learning pathways are more variable, and factors like L1 transfer or incidental exposure may have a stronger impact on early vocabulary acquisition than more systematic learning at higher levels, which could explain the slightly lower correlation for beginners (0.72).

From a pedagogical standpoint, these results support instructors' expert judgment in selecting and prioritizing technical terms, especially for advanced learners. For high-proficiency learners, the correlation is 0.813, indicating that experienced teachers can accurately identify which specialised terminology students are likely to know or need to acquire. However, when working with lower-proficiency learners, the slightly reduced alignment suggests that teachers may benefit from further diagnostic examinations to better assess students' vocabulary knowledge. These findings support and expand on earlier research by Dang et al. (2022b) and He and Godfroid (2019), demonstrating that teacher intuitions often mirror real learning patterns.

Overall, the correlation patterns demonstrate that teachers can more effectively adapt their lessons to students' developmental stages when they possess a thorough understanding of the vocabulary acquisition process. According to these findings, while frequency-informed word lists are valuable, curriculum designers should tailor them to suit learners at different proficiency levels. The findings provide empirical support for the notion that teacher competence aligns with actual language learning outcomes across developmental stages, thereby helping to bridge the gap between teacher cognition research and vocabulary studies.

7.10 Contributions of the Study

This section focuses on the methodological and theoretical contributions of the current thesis, which are outlined and discussed in the following sub-sections.

7.10.1 Methodological Contribution

This section discusses the methodological and theoretical contributions of the current research. The study offers three key methodological contributions to the field of ESP and vocabulary studies: (1) the development of an innovative genre-based approach to specialised corpus design for engineering discourse; (2) a refined framework for identifying technical vocabulary specific to the discipline; and (3) the creation and validation of a Yes/No test format to evaluate students' receptive knowledge of technical vocabulary.

The first methodological contribution lies in the development of a specialised corpus derived from Saudi engineering masters' dissertations. This corpus was constructed using a genre-oriented approach, where a genre refers to 'language use in a conventionalized communicative setting to express specific communicative goals within a disciplinary or social institution' (Bhatia, 2004: 23). Genre analysis is a method used to study specialised vocabulary in various types of discourse (Szudarski, 2018). As noted by Bowles (2012), there is a wealth of genre-oriented research into academic, business, and professional communication, yet no previous corpus-based research on engineering vocabulary has applied this genre-oriented approach. Therefore, the corpus design used in this study provides a valuable methodological contribution. Specifically, the specialised corpus has revealed the structural identity and textual nature of masters' dissertations as a particular academic genre, which is significant in the field of EAP vocabulary research (Thompson, 2005). Research in EAP has gradually shifted from register analysis to discourse or genre analysis, focusing on patterns, rhetorical structures, and phraseology (Thompson & Hunston, 2019).

The second methodological contribution is the development of an adaptable engineering technical word list suitable for improving EFL undergraduate learners' productive and receptive knowledge. The list contains pedagogical useful ETV both single-word items and multiword units. These word lists can serve as a guide for teachers to set both short- and long-term learning goals (Nation, 2013). At the beginning of a specialised course, both teachers and learners can set goals based on learners' linguistic backgrounds and existing general vocabulary levels. By dividing the word list into different levels, the learners' prior vocabulary knowledge is taken into account. This approach to adapting word lists to students' levels has been used by Dang et al. (2017) for the Academic Spoken Word List and by Lu (2018) for technical vocabulary in Traditional Chinese Medicine. In essence, the engineering technical word list in this study provides a model for making corpus-based word lists more suitable for ESP vocabulary learning and teaching.

The third methodological contribution of this study is the development and application of the Yes/No receptive test format to assess students' receptive knowledge of technical vocabulary. This thesis extends the traditional use of the Yes/No test beyond measuring general or academic vocabulary knowledge. While the Yes/No test format is well-established in vocabulary research, having been widely used in studies like those of Dang et al. (2022) to assess general word lists, this study breaks new ground by demonstrating its effectiveness for assessing technical vocabulary in engineering contexts. This adaptation required careful methodological considerations to address the unique characteristics of technical terminology, including the development of appropriate distractor items that account for potential partial knowledge of technical terms and the validation of test items with subject matter experts.

Using this assessment instrument alongside teacher evaluations to develop a more comprehensive evaluation framework for technical word lists is innovative, as it goes beyond simply adapting an established approach to a new field. By comparing students' self-reported knowledge from the Yes/No test with teachers' assessments of word utility, this dual-method approach offers several points of validation and sheds light on the connection between vocabulary selection and actual acquisition. By successfully applying this approach to technical vocabulary assessment, new avenues for ESP research are made possible, suggesting that this dependable and effective testing format can be effectively expanded beyond its conventional applications to specialised lexical domains.

Additionally, the study shows how this modified Yes/No test can serve both as a practical teaching tool to diagnose students' receptive knowledge of discipline-specific terms and as a research instrument for examining the characteristics of technical vocabulary knowledge. For technical vocabulary, where partial knowledge is common and pedagogically significant, the test's capacity to measure partial vocabulary knowledge – where students may recognize a term but not completely grasp its meaning or application – proves especially beneficial. Thus, this methodological development offers a model that might be applied to other specialised fields beyond engineering, contributing to both research practices in vocabulary studies and classroom evaluation in ESP contexts.

7.10.2 Theoretical Contributions

The present thesis makes important theoretical contributions to vocabulary research through the application of multiple conceptual frameworks. First, it advances our understanding of the characteristics of technical vocabulary in masters' dissertations, an academically crucial yet underexplored genre in engineering discourse. The research reveals that technical vocabulary in these dissertations exhibits complex lexical patterns, appearing across all frequency bands of Nation's (2012) BNC/COCA. Two key lists were developed: one for single-word items and another for multiword units. Importantly, the findings reveal that technical vocabulary permeates all vocabulary categories, including general and academic word lists, and is distributed across various discourse types within the dissertations.

From a semantic perspective, this thesis yields important insights into the meaning relationships of ETV items. The results reveal that many ETV items are polysemous, carrying discipline-specific meanings alongside more general interpretations, depending on contextual usage. However, this thesis identifies a subset of highly specialised terms that maintain singular, field-specific meanings regardless of discourse context. The nuanced understanding of technical vocabulary semantics contributes to ongoing theoretical debates surrounding the distinctions between general, academic, and technical vocabulary in specialised discourse.

Notably, this research is one of the first attempt to analyse technical vocabulary in engineering dissertations through a corpus-based approach, complemented by semantic rating and input from both learners and teachers. As discussed earlier, involving end users in word list research – both learners and teachers – has become a common practice, though it has been primarily applied to general vocabulary. This study successfully extends this approach to the analysis of technical vocabulary, re-contextualizing it within a specialised field. Recent studies have significantly advanced our understanding of the role played by technical vocabulary in specialised texts. While various researchers have defined and approached technical vocabulary differently, they consistently show that such vocabulary constitutes a significant portion of technical texts (Chung & Nation, 2003; Coxhead, 2018; Quero, 2015). In the context of ESP, examining the distribution of technical vocabulary across macrostructures helps illuminate the linguistic characteristics of texts within this genre or discourse.

Another key contribution of this thesis is the development of an MWU engineering technical vocabulary list. This adds to the body of literature in ESP on the functions of MWUs as a useful learning resource for constructing and comprehending specialised discourse. These findings suggest that certain technical terms are most meaningful when considered as part of MWUs rather than in isolation as single words. The thesis found a substantial number of MWUs in engineering technical vocabulary, using a root structure with at least one single-word identified ETV. This contributes to the growing body of research on MWUs in ESP, as seen in studies such as Lu (2018) in Traditional Chinese Medicine; Benson and Coxhead (2022) and Benson (2020) in rugby; Ha (2015) in finance; Coxhead et al. (2017) in university tutorials and laboratories; and Ward (2007) in engineering. MWUs are now widely recognised as an essential part of the vocabulary learning curriculum, alongside single-word items (Pellicer-Sánchez, 2020). Therefore, the list developed in this study represents a valuable resource for EFL engineering students. As EFL learners often acquire vocabulary through single-word lists, the integration of technical MWUs provides a richer and more contextually accurate representation of domain-specific language use (Benson, 2020).

7.11 Pedagogical Implications of the Study

The present thesis has significant pedagogical implications for vocabulary development in ESP. The pedagogical useful ETV lists, both the single-word and multiword units developed in this thesis, will be helpful for university engineering students, enabling them to enhance their knowledge of engineering technical vocabulary. The engineering technical word list generated here offers both students and teachers a valuable pool of lexical resources to focus on. However, it is important to emphasize that working with this word list should not involve rote memorization (Lu, 2018; Dang, 2017). As previous research highlights (Coxhead, 2000), these items should not be learned or taught in isolation but rather integrated with students' existing linguistic knowledge and the contextual characteristics of the words themselves.

Throughout this thesis, I have emphasized the importance of corpus analysis in understanding and describing vocabulary use. Such analyses not only deepen our comprehension of vocabulary but also have practical implications for language education. For example, frequency-based analyses can guide teachers and materials developers in prioritizing which vocabulary items to include in language instruction and textbooks.

The single-word list and multiword unit list developed in this study have practical applications for data-driven learning and for creating teaching materials in ESP and specialised engineering discourse. Nesi (2013) emphasizes the role of ESP corpora in both data-driven learning and the development of instructional materials, a point supported by Römer (2011) and Flowerdew (2009), who distinguish between the direct and indirect applications of corpus linguistics in language teaching. As Gavioli (2005, p. 69) notes, 'concordances provide a way to look at typical (or atypical), conventional (or non-conventional) uses of language'.

The findings of this study have significant implications for teaching and learning technical vocabulary in English-medium engineering programmes. The engineering technical word lists developed from the EMDC provide a structured guide for teachers to set both long-term and short-term learning objectives (Nation, 2013). At the outset of specialised courses, teachers and learners can tailor these goals according to students' existing general vocabulary knowledge and linguistic backgrounds. The lexical profiling of the EMDC, mapped against Nation's (2012) BNC/COCA lists and Schmitt and Schmitt's (2014) high-, mid-, and low-frequency bands, demonstrates that the corpus is lexically demanding. Achieving 95–98% coverage requires knowledge of 5,000–7,000-word families, including supplementary and engineering-specific lists. High-frequency words form the foundation of academic cohesion, supplementary lists contribute 6.7% of coverage, and engineering-specific terms, though limited (0.54%), are essential for conveying discipline-specific meaning. These insights are relevant for both undergraduate reading activities and graduate engagement with engineering dissertations.

The study further highlights the distribution of technical single-word and multiword units in the corpus. Semantic rating analysis of 570 engineering technical single words revealed that 61.3% are high-frequency, 25.5% mid-frequency, 4.1% low-frequency, 2% from supplementary lists, and 7.1% from the engineering base word list. Understanding this distribution across dissertation sections provides practical guidance for teaching, indicating how technical vocabulary is represented within the engineering genre. Teachers can use these lists to design targeted lessons, while learners can systematically expand their knowledge of specialised vocabulary.

Knowledge of technical vocabulary is critical for EFL engineering learners in EMI contexts. The frequency-based lists offer a valuable pool of lexical resources for focused learning. However, acquiring this vocabulary should go beyond rote memorization (Dang, 2017) or decontextualized learning (Coxhead, 2000). Learners benefit from engaging with technical terms in context, integrating them into academic reading, writing, and discussion, and linking new terms to prior knowledge. Teachers should first provide comprehensible, meaning-focused input through reading and listening (Nation, 2013), before explicitly drawing attention to target vocabulary, ensuring learners encounter technical terms both contextually and analytically to support deeper understanding and retention.

The technical word lists also inform the design of assessments. Test developers can select items aligned with instructional activities and learning goals (Malmström, Pecorari, & Shaw, 2018). For example, if instruction emphasizes receptive knowledge, assessments should target this dimension. The lists can also guide syllabus design or the enhancement of existing courses. Once learning objectives are established, the engineering technical word lists can underpin a specialised lexical syllabus, enabling teachers to create activities that help learners acquire, consolidate, and apply technical vocabulary meaningfully within engineering contexts.

Additionally, teacher ratings of technical vocabulary underscore its role in improving learners' comprehension of engineering texts. The lists of ETV items generated in this study can assist teachers, textbook writers, curriculum developers, and test designers in creating learning goals, instructional materials, and activities that focus on the most relevant technical words at different stages of L2 vocabulary development.

Given the lexically demanding nature of the engineering field, fostering autonomous vocabulary learning strategies is particularly beneficial. Strategies such as inferring meaning from context, recognizing lexical patterns, analyzing nominal compounds, searching for synonyms, and understanding root words and affixes enable students to decode and retain complex technical terms (Williams, 1985). For example, combining roots such as *therm(o)* (heat) with *-dynamics* forms *thermodynamics*, while *electro-* and *-magnetic* produce *electromagnetic*. Recognizing familiar components allows learners to deconstruct unfamiliar terms, verify contextual predictions, and enhance retention through meaningful connections between related concepts (e.g., *thermocouple*, *thermoplastic*, *thermodynamics*). Explicit instruction on root structures strengthens students' ability to acquire and apply discipline-specific language, ultimately improving comprehension and communication in engineering contexts (Nation, 2013).

Overall, the findings highlight that the single-word and multiword technical vocabulary lists provide a practical foundation for supporting L2 engineering learners. These lists can guide systematic vocabulary acquisition, inform teaching and assessment, and help learners engage effectively with specialised academic discourse. They serve as a valuable resource for students, teachers, course designers, and material developers, offering a structured starting point for learning technical vocabulary. However, mastery of high-frequency items on the lists should be complemented with strategies to acquire lower-frequency technical terms, alongside a strong foundation in general high-frequency vocabulary.

7.12 Limitations of the Study

This thesis has several limitations. The first limitation pertains to corpus compilation. The EMDC consisted of a substantial amount of data derived from masters' dissertations written by Saudi engineering students. However, the corpus size is relatively small by contemporary standards, which should be acknowledged as a weakness. While efforts were made to ensure a representative selection of texts, it is possible that a different or broader selection of textbooks and journal articles could have yielded different results. Therefore, the corpus used in this research may not fully capture all linguistic patterns in the specialised discourse of engineering, nor represent them in precisely the correct proportions.

The second limitation is related to the corpus content. This study concentrated solely on masters' dissertations as the primary genre, excluding other important genres within the field of engineering, such as textbooks and journal articles, which students frequently engage with to meet their lexical needs. While dissertations are a key academic output for students, textbooks and journal articles are crucial reference materials. This research should therefore be viewed as an initial step in the investigation of engineering vocabulary, offering a foundation and direction for future studies, as outlined in Section 7.14.

Another limitation is related to the initial scope of the study. The initial overarching purpose for conducting this research was to investigate engineering dissertations produced by native and non-native students. However, it became evident that such an undertaking required distinct analytical approaches and considerable time to conduct both quantitative and semantic analyses comprehensively. Given these constraints, the study had to adopt a more focused approach, prioritizing technical vocabulary to ensure a thorough and meaningful analysis. While the initial plan aimed to encompass both vocabulary types, a thorough analysis of both within the available timeframe proved unfeasible.

Moreover, the use of Nation's (2012) BNC/COCA word lists in the lexical profiling and vocabulary load analysis posed another limitation. This research utilized the abbreviation list (BASEWRD 34) from the BNC/COCA lists, which was not an ideal option for specialised engineering vocabulary. A more accurate approach would have been to create an abbreviation list specifically for engineering at the outset of the study. Some abbreviations in the BNC/COCA list were actually technical words in this specialised corpus, and many technical abbreviations, which are highly specialised, did not appear in general English texts. For more precise results, future ESP research should consider creating new abbreviation lists tailored to specific fields, as Coxhead and Demecheleer (2018) did for their plumbing corpus.

Another limitation pertains to practical application, specifically the relatively small number of participants in the Yes/No test and teacher rating of word usefulness. The number of teachers involved in rating the selected words was also limited. Additionally, only 210 out of 930 technical words were included in the Yes/No and teacher rating tasks, unlike in the study conducted by Dang (2017), which involved the entire vocabulary list. This selective sampling approach presents a limitation in the generalizability of the results, particularly concerning students' receptive knowledge and the word usefulness ratings in Study 3. As a result, the findings should be interpreted with caution and regarded as tentative rather than definitive.

7.13 Concluding Remarks

The present thesis marks a significant first step in using insights from both teachers and learners to use corpus-based data in evaluating technical vocabulary in engineering discourse. While a few recent studies have applied similar approaches with high-frequency word lists, such as the work of Dang, et al., (2022a), this thesis expands the scope by focusing on technical vocabulary in the specific context of engineering masters' dissertations. It builds upon previous research, including Dang and Webb's (2016) study, which compared general high-frequency word lists like Nation's (2012) BNC/COCA 2000 and Brezina and Gablasova's (2015) new-GSL, but with a focus on L2 learners' vocabulary knowledge and teacher perceptions of word usefulness.

Addressing the gap in research on technical vocabulary within the dissertation and thesis genres, this thesis conducted three interrelated studies to explore the nature of vocabulary in engineering masters' dissertations written by non-native English speakers. The first two studies identified the most appropriate general high-frequency word list for these learners, considering insights from corpus linguistics and input from teachers and learners.

A key outcome of this research was the development of an engineering technical vocabulary word list, tailored to the proficiency levels of L2 learners, and used to investigate the technical vocabulary present in engineering dissertations as a specialised genre of academic discourse. Particularly, the study examines engineering technical vocabulary from the perspectives of both single-word and multiword units, an area that, to the best of the researcher's knowledge, has not been previously explored within the engineering discipline. Additionally, the research focuses specifically on masters' dissertations, a genre that existing evidence suggests remains under investigated in ESP research and academic discourse studies. The integration of corpus-based and sematic approach contribute to both methodological and theoretical advances in corpus-based research. Moreover, the findings highlight the importance of technical vocabulary in the teaching and learning of engineering.

In conclusion, the overall analysis reveals that the EMDC has a very distinct lexical profile compared to general written English. The EMDC texts contained a smaller proportion of high-frequency vocabulary, with a greater emphasis on mid- and low-frequency vocabulary, along with items from supplementary word lists. Additionally, a wide range of vocabulary lay outside these established word lists, suggesting that the vocabulary demands of the EMDC are particularly challenging.

7.14 Recommendations for Future Research

This section offers recommendations for future research based on the issues that emerged while addressing the research questions of this study. These new findings provide a foundation for guiding future studies and expanding upon the current work. One area for future research is the exploration of technical vocabulary beyond the focus of this thesis, which was limited to engineering masters' dissertations as a genre of engineering discourse. While these dissertations provide valuable insights into specialised language, other types of texts such as textbooks, journal articles, tutorials, and conference presentations also play a crucial role in the development of disciplinary literacy. These materials deserve further examination, particularly in the context of engineering, to offer a more comprehensive view of specialised language use.

Another important area to consider in future studies is the analysis of technical vocabulary in spoken discourse across engineering and other disciplines. Investigating spoken language in professional and academic settings would enhance our understanding of how specialised vocabulary is used in real-world communication. This could provide valuable insights for researchers, ESP teachers, and learners, contributing to a more holistic understanding of vocabulary in these specialised fields.

Future studies could adopt a longitudinal approach, tracking students' vocabulary development over an extended period, such as two years, and also students' lexical knowledge at different levels of mastery in terms of different aspects of lexical competence. Such research would offer a deeper understanding of how students' receptive and productive vocabulary evolves and whether they continue to face similar challenges over time. This would help in identifying persistent issues in vocabulary acquisition and inform strategies to address them effectively and also inform the theory of L2 vocabulary learning.

Finally, my thesis has contributed to the broad field of applied linguistics by developing ETV lists of both single-word items and multiword units through combining corpus-based and semantic methods, using masters' dissertations as an essential academic genre.

References

Ackermann, K., & Chen, Y.-H. (2013). Developing the Academic Collocation List (ACL) – A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, *12*(4), 235–247. https://doi.org/10.1016/j.jeap.2013.08.002

Akbarian, I. (2010). The relationship between vocabulary size and depth for ESP/EAP learners. *System (Linköping)*, *38*(3), 391–401. https://doi.org/10.1016/j.system.2010.06.013

Al Asmari, A., & Malik, A. G. (2016). The use of EFL authentic materials: A gender-line university teachers' perspective. International Journal of Applied Linguistics & English Literature, 5(2), 128–138. https://doi.org/10.7575/aiac.ijalel.v.5n.2p.128

Al Roomy, M., & Alhawsawi, S. (2019). Understanding reading strategies of EFL Saudi students. English Language Teaching, 12(6), 33–44. https://doi.org/10.5539/elt.v12n6p33

Al-Khresheh, M. H., & Al-Ruwaili, S. F. (2020). An exploratory study into vocabulary learning strategies used by Saudi EFL learners. Journal of History Culture and Art Research, 9(2), 288–302.

Al-Seghayer, K. (2021). Factors underlying current Saudi EFL teachers' approaches to teaching the four macro and micro language skills. International Journal of English Linguistics, 11(5), 44–61. https://doi.org/10.5539/ijel.v11n5p44

Alabdulaziz, M. S. R. (2019). Overview of the education system in the Kingdom of Saudi Arabia. *International Journal of Information Technology* (IJIT), 5(2), 1–12. https://www.ijitjournal.org/volume-5/issue-2/IJIT-V5I2P1.pdf

Alanazi, K., & Curle, S. (2025). Overcoming linguistic hurdles: Challenges and strategies for English-medium instruction in Saudi Arabian medical education. Social Sciences & Humanities Open, 11, 101334. https://doi.org/10.1016/j.ssaho.2025.101334

Alfahadi, A. (2012). *Saudi teachers' views on appropriate cultural models for EFL textbooks: Insights into TESOL teachers' management of global cultural flows and local realities in their teaching worlds* (Publication No. U590625). [Doctoral dissertation, University of Exeter]. ProQuest Dissertations & Theses.

Alfehaid, A. (2018). Using English as a medium of instruction in a Saudi university: Experiences and implications. Asian EFL Journal, 20(12), 60–101.

Alhamami, M. (2021). English as a medium of instruction (EMI) in computing undergraduate programs. Education and Information Technologies, 26(5), 6549–6562.

Alharbi, S. H. (2021). The struggling English language learners: Case studies of English language learning difficulties in EFL context. *English Language Teaching, 14*(11), 108–117. https://eric.ed.gov/?id=EJ1318092

Alrabai, F. (2018). Learning English in Saudi Arabia. In C. Moskovsky & M. Picard (Eds.), English as a foreign language in Saudi Arabia (pp. 102–119). Routledge.

Alsaif, A., & Milton, J. (2012). Vocabulary input from school textbooks as a potential contributor to the small vocabulary uptake gained by English as a foreign language learners in Saudi Arabia. *Language Learning Journal*, *40*(1), 21–33. https://doi.org/10.1080/09571736.2012.658221

Alshahrani, A. (2020). Integrating education for sustainability into English as a foreign language instruction in the Kingdom of Saudi Arabia (Doctoral dissertation, Curtin University).

Anthony, L. (2014). *AntWordProfiler* (Version 1.4.1) [Computer software]. Waseda University. www.laurenceanthony.net/software

Anthony, L. (2020*). AntConc* (3.5.8) [Computer software]. Waseda University. https://www.laurenceanthony.net

Anthony, L. (2022). What can corpus software do? In A. O'Keefe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 103–125). Routledge. https://doi.org/10.4324/9780203856949-14

Bardel, C., Gudmundson, A., & Lindqvist, C. (2012). Aspects of lexical sophistication in advanced learners' oral production. Vocabulary acquisition and use in L2 French and Italian. *Studies in Second Language Acquisition*, *34*(2), 269–290. https://doi.org/10.1017/S0272263112000058

Barnawi, O. Z., & Al-Hawsawi, S. (2017). English language education policy in the Kingdom of Saudi Arabia: Current trends, issues and challenges. In R. Kirkpatrick (Ed.), English language education policy in the Middle East and North Africa (pp. 199–222). Springer.

Barnbrook, G. (2019). *Language and computers: A practical introduction to the computer analysis of language*. Edinburgh University Press. https://doi.org/10.1515/9781474464185

Basturkmen, H. (2006). *Ideas and options in English for specific purposes*. Routledge. https://doi.org/10.4324/9781410617040

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, *6*(4), 253–279. https://doi.org/10.1093/ijl/6.4.253

Beasley, W. H., & Rodgers, J. L., (2009). Resampling methods. In W. H. Beasley & J. L. Rodgers (Eds.), *The SAGE handbook of quantitative methods in psychology* (pp. 362–386). SAGE Publications. https://doi.org/10.4135/9780857020994.n16

Beeckmans, R., Eyckmans, J., Janssens, V., Dufranne, M., & van de Velde, H. (2001). Examining the yes/no vocabulary test: Some methodological issues in theory and practice. *Language Testing*, *18*(3), 235–274. https://doi.org/10.1191/026553201680188960

Benson, M. J., & Basturkmen, H. (2006). *Teaching the English for academic purposes course: A practical guide*. University of Michigan Press.

Benson, S. (2020). *Dot the pill down: Investigating the linguistic needs of foreign rugby players and lexicon of spoken rugby discourse* [Unpublished doctoral dissertation]. Te Herenga Waka – Victoria University of Wellington.

Benson, S., & Coxhead, A. (2022). Technical single and multiword unit vocabulary in spoken rugby discourse. *English for Specific Purposes*, *66*, 111–130. https://doi.org/10.1016/j.esp.2022.02.001

Bhatia, V. K. (2002). A generic view of academic discourse. (2002). In J. Flowerdew (Ed.), *Academic discourse* (pp. 31–49). Routledge. https://doi.org/10.4324/9781315838069-9

Bhatia, V. (2004). *Worlds of written discourse: A genre-based view*. Bloomsbury Publishing.

Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing, 8*(4), 243–257. https://doi.org/10.1093/llc/8.4.243

Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes*, *5*(2), 97–116. https://doi.org/10.1016/j.jeap.2006.05.001

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, *14*(3), 275–311. https://doi.org/10.1075/ijcl.14.3.08bib

Biber, D. (2019). Text-linguistic approaches to register variation. *Register Studies, 1*(1), 42–75. https://doi.org/10.1075/rs.18007.bib

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, *26*(3), 263–286. https://doi.org/10.1016/j.esp.2006.08.003

Biber, D., & Conrad, S. (2009). *Register, genre, and style.* Cambridge University Press.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at …: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, *25*(3), 371–405. https://doi.org/10.1093/applin/25.3.371

Biber, D., Conrad, S., & Reppen, R. (1994). Corpus-based approaches to issues in applied linguistics. *Applied Linguistics*, *15*(2), 169–189. https://doi.org/10.1093/applin/15.2.169

Biber, D., & Reppen, R. (2002). What does frequency have to do with grammar teaching? Studies in Second Language Acquisition, 24(2), 199–208. https://doi:10.1017/S0272263102002048

Bitchener, J. (2009). *Writing an applied linguistics thesis or dissertation: A guide to presenting empirical research*. Bloomsbury Publishing.

Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, *36*(2), 81–109. https://doi.org/10.1017/S0261444803001903

Borg, S. (2006). The distinctive characteristics of foreign language teachers. *Language Teaching Research*, *10*(1), 3–31. https://doi.org/10.1191/1362168806lr182oa

Bowles, H. (2012). Analyzing languages for specific purposes discourse. *The Modern Language Journal, 96*, 43–58. http://www.jstor.org/stable/41478790

Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press. https://doi.org/10.1017/9781316410899

Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics*, *36*(1), 1–22. https://doi.org/10.1093/applin/amt018

Brezina, V., Hawtin, A., & McEnery, T. (2021). The Written British National Corpus 2014 – design and comparability. *Text & Talk*, *41*(5), 595–615. https://doi.org/10.1515/text-2020-0052

British Educational Research Association (BERA) (2018). Ethical guidelines for educational research. British Educational Research Association. Retrieved January 29, 2021, from https://www.bera.ac.uk/resources/all-publications/resources-for-researchers

Brown, J. D., & Rodgers, T. S. (2002). *Doing second language research: An introduction to the theory and practice of second language research for graduate/master's students in TESOL and applied linguistics, and others*. Oxford University Press.

Browne, C. (2014). A New General Service List: The better mousetrap we've been looking for? *Vocabulary Learning and Instruction, 3*(2), 1–10. https://doi.org/10.7820/vli.v03.2.browne

Browne, C., Culligan, B., & Phillips, J. (2013). A new academic word list. Retrieved from http://www.newacademicwordlist.org/

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Bunton, D. R. (1998). Linguistic and textual problems in Ph.D. and M. Phil theses: An analysis of genre moves and metatext [Doctoral dissertation]. HKU Theses Online (HKUTO). https://hub.hku.hk/bitstream/10722/39580/1/FullText.pdf

Byrd, P., & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL, 5*(5), 31–64.

Cailing, L., Boers, F., & Coxhead, A. (2021). Exploring learners' understanding of technical vocabulary in traditional Chinese medicine. *Studies in Second Language Learning and Teaching*, *11*(1), 71–101. https://doi.org/10.14746/ssllt.2021.11.1.4

Cameron, L. (2002). Measuring vocabulary size in English as an additional language. *Language Teaching Research*, *6*(2), 145–173. https://doi.org/10.1191/1362168802lr103oa

Chang, C.-F., & Kuo, C.-H. (2011). A corpus-based approach to online materials development for writing research articles. *English for Specific Purposes*, *30*(3), 222–234. https://doi.org/10.1016/j.esp.2011.04.001

Charles, M. (2003). 'This mystery …': A corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines. *Journal of English for Academic Purposes*, *2*(4), 313–326. https://doi.org/10.1016/S1475-1585(03)00048-1

Charles, M. (2006). The construction of stance in reporting clauses: A cross-disciplinary study of theses. *Applied Linguistics*, *27*(3), 492–518. https://doi.org/10.1093/applin/aml021

Chen, C.-M., & Li, Y.-L. (2010). Personalised context-aware ubiquitous learning system for supporting effective English vocabulary learning. *Interactive Learning Environments*, *18*(4), 341–364. https://doi.org/10.1080/10494820802602329

Chen, Q., & Ge, G. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *English for Specific Purposes*, *26*(4), 502–514. https://doi.org/10.1016/j.esp.2007.04.003

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, *14*(2), 30–49. http://dx.doi.org/10125/44213

Cheng, W. (2012). *Exploring corpus linguistics: Language in action*. Routledge. https://doi.org/10.4324/9780203802632

Chung, T. M. (2003). A corpus comparison approach for terminology extraction. *Terminology. International Journal of Theoretical and Applied Issues in Specialised Communication, 9(*2), 221–246. https://doi.org/10.1075/term.9.2.05chu

Chung, T. M., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language*, *15*(2), 103–116.

Chung, T. M., & Nation, P. (2004). Identifying technical vocabulary. *System (Linköping)*, *32*(2), 251–263. https://doi.org/10.1016/j.system.2003.11.008

Cohen, L., Manion, L., & Morrison, K. (2018). The ethics of educational and social research. In L. Choen, L. Manion & K. Morrison (Eds.), *Research methods in education* (8th ed., pp. 111–143). Routledge. https://doi.org/10.4324/9781315456539-7

Connor, U. & Upton, T. A. (Eds.) (2004). *Applied corpus linguistics. A multi-dimensional perspective.* Rodopi.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, *23*(4), 397–423. https://doi.org/10.1016/j.esp.2003.12.001

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–238. https://doi.org/10.2307/3587951

Coxhead, A. (2011). The academic word list 10 years on: Research and teaching implications. *TESOL Quarterly*, *45*(2), 355–362. https://doi.org/10.5054/tq.2011.254528

Coxhead, A. (2013). Vocabulary and ESP. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 115–132). Wiley-Blackwell.

Coxhead, A. (2016). Reflecting on Coxhead (2000), "A new academic word list." *TESOL Quarterly*, *50*(1), 181–185. https://doi.org/10.1002/tesq.287

Coxhead, A. (2018). *Vocabulary and English for specific purposes research*. London: Routledge.

Coxhead, A. (2020). Academic vocabulary. In S. Webb (Ed), *The Routledge handbook of vocabulary studies* (pp. 97–110). Routledge.

Coxhead, A. (2021). Vocabulary in English in tertiary contexts: Connecting research and learning. *LEARN Journal: Language Education and Acquisition Research Network*, *14*(1), 1–14.

Coxhead, A., & Boutorwick, T. J. (2018). Longitudinal vocabulary development in an EMI international school context: Learners and texts in EAL, maths, and science. *TESOL Quarterly*, *52*(3), 588–610. https://doi.org/10.1002/tesq.450

Coxhead, A., Dang, T. N., & Mukai, S. (2017). Single and multi-word unit vocabulary in university tutorials and laboratories: Evidence from corpora and textbooks. *Journal of English for Academic Purposes*, *30*, 66–78. https://doi.org/10.1016/j.jeap.2017.11.001

Coxhead, A., & Demecheleer, M. (2018). Investigating the technical vocabulary of Plumbing. *English for Specific Purposes*, *51*, 84–97. https://doi.org/10.1016/j.esp.2018.03.006

Coxhead, A., Demecheleer, M., & McLaughlin, E. (2016). The technical vocabulary of carpentry: Loads, lists and bearings. *TESOLANZ Journal, 24,* 38–71.

Coxhead, A., & Hirsch, D. (2007). A pilot science-specific word list. *Revue Française de Linguistique Appliquée, 12*(2), 65–78.

Coxhead, A., McLaughlin, E., & Reid, A. (2019). The development and application of a specialised word list: The case of fabrication. *Journal of Vocational Education & Training*, *71*(2), 175–200. https://doi.org/10.1080/13636820.2018.1471094

Coxhead, A., & Nation, P. (2001). The specialised vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 252–267). Cambridge University Press. https://doi.org/10.1017/CBO9781139524766.020

Coxhead, A., Rahmat, Y., & Yang, L. (2020). Academic single and multiword vocabulary in EFL textbooks: Case studies from Indonesia and China. *TESOLANZ Journal, 28*, 75–88.

Csomay, E., & Petrović, M. (2012). "Yes, your honor!": A corpus-based study of technical vocabulary in discipline-related movies and TV shows. *System (Linköping)*, *40*(2), 305–315. https://doi.org/10.1016/j.system.2012.05.004

Csomay, E., & Prades, A. (2018). Academic vocabulary in ESL student papers: A corpus-based study. *Journal of English for Academic Purposes*, *33*, 100–118. https://doi.org/10.1016/j.jeap.2018.02.003

Culligan, B. (2015). A comparison of three test formats to assess word difficulty. *Language Testing*, *32*(4), 503–520. https://doi.org/10.1177/0265532215572268

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

Dang, T. (2020). Vietnamese non-English major EFL university students' receptive knowledge of the most frequent English words. *VNU Journal of Foreign Studies, 36*(3). https://doi.org/10.25073/2525-2445/vnufs.4553

Dang, T. N. Y. (2018a). A hard science spoken word list. *ITL – International Journal of Applied Linguistics*, *169*(1), 44–71. https://doi.org/10.1075/itl.00006.dan

Dang, T. N. Y. (2018b). The nature of vocabulary in academic speech of hard and soft-sciences. *English for Specific Purposes*, *51*, 69–83.

Dang, T. N. Y. (2022). Vocabulary in academic lectures. *Journal of English for Academic Purposes*, *58*, 101123. https://doi.org/10.1016/j.jeap.2022.101123

Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning*, *67*(4), 959–997. https://doi.org/10.1111/lang.12253

Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, *33*(1), 66–76. https://doi.org/10.1016/j.esp.2013.08.001

Dang, T. N. Y., & Webb, S. (2016). Making an essential word list for beginners. In I. S. P. Nation (Ed.), *Making and using word lists for language learning and testing* (pp. 153–167). John Benjamins Publishing. https://doi.org/10.1075/z.208.c15

Dang, T. N. Y., Webb, S., & Coxhead, A. (2022a). Evaluating lists of high-frequency words: Teachers' and learners' perspectives. *Language Teaching Research*, *26*(4), 617–641. https://doi.org/10.1177/1362168820911189

Dang, T. N. Y., Webb, S., & Coxhead, A. (2022b). Relationships between lexical coverage, learner knowledge, and teacher perceptions of the usefulness of high-frequency words. *Foreign Language Annals*, *55*(4), 1212–1230. https://doi.org/10.1111/flan.12663

Dang, Y. (2017). *Investigating vocabulary in academic spoken English: Corpora, teachers, and learners* (Doctoral dissertation). Te Herenga Waka – Victoria University of Wellington. https://openaccess.wgtn.ac.nz/articles/thesis/Investigating_vocabulary_in_academic_spoken_English_Corpora_teachers_and_learners/17060051?file=31547354

De Jong, F. (2018). English as a medium of instruction in higher education in a cross-national context (Unpublished doctoral dissertation). University of Groningen.

https://www.rug.nl/research/portal/en/publications/english-as-medium-of-instruction-in-higher-education-in-a-crossnational-context(7f276a49-28eb-4efa-ac77-69cd2a7ade27).html

Dörnyei, Z., & Taguchi, T. (2009). *Questionnaires in second language research: Construction, administration, and processing.* (2nd ed.). Taylor & Francis Group.

Dudley-Evans, T. (1999). An overview of ESP in the 1990s. In T. Dudley-Evans & M. J. St John (Eds.), *Developments in English for specific purposes: A multi-disciplinary approach* (pp. 1–13). Cambridge University Press.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, *19*(1), 61–74.

Đurović, Z. (2021). Corpus linguistics methods for building ESP word lists, glossaries and dictionaries on the example of a marine engineering word list. *Lexikos*, *31*, 259–282. https://doi.org/10.5788/31-1-1647

Durovic, Z., Vukovic-Stamatovic, M., & Vukicevic, M. (2021). How much and what kind of vocabulary do marine engineers need for adequate comprehension of ship instruction books and manuals? *Círculo de Lingüística Aplicada a La Comunicación*, *88*, 123–133. https://doi.org/10.5209/clac.78300

Durrant, P. (2014). Discipline and level specificity in university students' written vocabulary. *Applied Linguistics*, *35*(3), 328–356. https://doi.org/10.1093/applin/amt016

Durrant, P. (2016). To what extent is the academic vocabulary list relevant to university student writing? *English for Specific Purposes*, *43*, 49–61. https://doi.org/10.1016/j.esp.2016.01.004

D'Angelo, L. (2016). The academic poster genre: Friend or foe? In K. Hyland (Ed.), *The Routledge handbook of English for academic purposes* (pp. 392–402). Routledge. https://doi.org/10.4324/9781315657455

Díaz-Negrillo, A., Thompson, P. & Ballier, N. (Eds.) (2013). *Automatic treatment and analysis of learner corpus data*. John Benjamins Publishing.

Egbert, J., & Gray, L. (2019). Introduction to the 'Handbook of technology and second language teaching and learning'. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. xx–xx). Wiley-Blackwell.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*(2), 143–188. https://doi.org/10.1017/s0272263102002024

Elyas, T. & Badawood, O. (2017). English language educational policy in Saudi Arabia post 21st century: Enacted curriculum, identity, and modernisation: A critical discourse analysis approach. In *FIRE: Forum for International Research in Education*, *3*(3), 70–81. http://preserve.lehigh.edu/fire/vol3/iss3/3

Eyckmans, J. (2004). *Measuring receptive vocabulary size. Reliability and validity of the yes/no vocabulary test for French-speaking learners of Dutch*. LOT.

Farrell, P. (1990). Vocabulary in ESP: A lexical analysis of the English of electronics and a study of semi-technical vocabulary (CLCS Occasional Paper No. 25). Trinity College, Centre for Language and Communication Studies.

Flowerdew, J., 1996. Concordancing in language learning. In M. Pennington (Ed.), *The power of CALL* (pp. 97–113). Athelstan.

Flowerdew, J. (2002). A pedagogic grammar of signalling nouns in discourse.

Flowerdew, J. (2009). Corpora in language teaching. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 327–350). Blackwell Publishing.

Fox, J., & Tigchelaar, M. (2015). Creating an engineering academic formulas list. *Journal of Teaching English for Specific and Academic Purposes, 3*(2), 295–304.

Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and usage*. Houghton Mifflin.

Fraser, S. (2005). The lexical characteristics of specialised texts. In K. Bradford-Watts, C. Ikeguchi & M. Swanson (Eds.), *JALT2004 conference proceedings* (pp. 318–327). JALT.

Fraser, S. (2009). Technical vocabulary and collocational behaviour in a specialised corpus. In A. Harris & A. Brandt (Eds.), *Language, learning and context: Proceedings of the 42nd annual meeting of the British Association for Applied Linguistics* (pp. 43–48). Scitsiugnil Press.

Freebody, P., & Anderson, R. C. (1983). Effects of vocabulary difficulty, text cohesion, and schema availability on reading comprehension. *Reading Research Quarterly*, *18*(3), 277–294. https://doi.org/10.2307/747389

Gaffas, Z. M. (2019). Students' perceptions of the impact of EGP and ESP courses on their English language development: Voices from Saudi Arabia. Journal of English for Academic Purposes, 42, 100797. https://doi.org/10.1016/j.jeap.2019.100797

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, *35*(3), 305–327. https://doi.org/10.1093/applin/amt015

Garnier, M., & Schmitt, N. (2015). The PHaVE list: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research*, *19*(6), 645–666. https://doi.org/10.1177/1362168814559798

Gavioli, L. (2005). *Exploring corpora for ESP learning*. John Benjamins Publishing.

Gay, L. R., Mills, G. E. & Airasian, P. (2006). *Educational research: Competencies for analysis and applications* (8th ed.). Pearson Education International.

Ghadessy, P. (1979). Frequency counts, words lists, and materials preparation: A new approach. *English Teaching Forum, 17*(1), 24–27.

Gillaerts, P., & Van de Velde, F. (2010). Interactional metadiscourse in research article abstracts. *Journal of English for Academic Purposes*, *9*(2), 128–139. https://doi.org/10.1016/j.jeap.2010.02.004

González Pueyo, I., & Val, S. (1996). The construction of technicality in the field of plastics: A functional approach towards teaching technical terminology. *English for Specific Purposes*, *15*(4), 251–278. https://doi.org/10.1016/S0889-4906(96)00011-7

Graham, K. M., & Eslami, Z. R. (2019). Attitudes toward EMI in East Asia and the Gulf: A systematic review. *Language Problems and Language Planning, 43*(1), 8–31. https://doi.org/10.1075/lplp.00021.gra

Granger, S., & Larsson, T. (2021). Is core vocabulary a friend or foe of academic writing? Single-word vs multi-word uses of thing. *Journal of English for Academic Purposes*, *52*, 100999. https://doi.org/10.1016/j.jeap.2021.100999

Gregori-Signes, C., Clavel-Arroitia, B., FernandezFuertes, R., PizarroSanchez, FuertesOlivera, P., LopezArroyo, B., SaezHidalgo, A., VelascoSacristan, M., Garcia, P., Amo, M., SastreRuano, M., & AlvarezDeLaFuente, E. (2015). Analysing lexical density and lexical diversity in university students' written discourse. *Procedia, Social and Behavioral Sciences*, *198*, 546–556. https://doi.org/10.1016/j.sbspro.2015.07.477

Ha, Y. H. (2015). Technical vocabulary in finance: A corpus-based study of annual reports and earnings calls [Doctoral dissertation]. HKU Theses Online (HKUTO). http://dx.doi.org/10.5353/th_b5570790

Ha, A. Y. H., & Hyland, K. (2017). What is technicality? A technicality analysis model for EAP vocabulary. *Journal of English for Academic Purposes*, *28*, 35–49. https://doi.org/10.1016/j.jeap.2017.06.003

Harrington, M., & Carey, M. (2009). The on-line yes/no test as a placement tool. *System (Linköping)*, *37*(4), 614–626. https://doi.org/10.1016/j.system.2009.09.006

Harsch, C., & Hartig, J. (2015). What are we aligning tests to when we report test alignment to the CEFR? *Language Assessment Quarterly*, *12*(4), 333–362. https://doi.org/10.1080/15434303.2015.1092545

Hatch, E. M., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Newbury House.

He, X., & Godfroid, A. (2019). Choosing words to teach: A novel method for vocabulary selection and its practical application. *TESOL Quarterly*, *53*(2), 348–371. https://doi.org/10.1002/tesq.483

Heatley, A., Nation, P., & Coxhead, A. (1994*)*. Range [Computer software]. Victoria University of Wellington. https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs

Hernández, M., Costa, A., & Arnon, I. (2016). More than words: Multiword frequency effects in non-native speakers. *Language, Cognition and Neuroscience*, *31*(6), 785–800. https://doi.org/10.1080/23273798.2016.1152389

Higginbotham, G., & Reid, J. (2019). The lexical sophistication of second language learners' academic essays. *Journal of English for Academic Purposes*, *37*, 127–140. https://doi.org/10.1016/j.jeap.2018.12.002

Hinkel, E. (2004). *Teaching academic ESL writing: Practical techniques in vocabulary and grammar*. L. Erlbaum Associates.

Hinkel, E. (Ed.). (2018). *Teaching essential units of language: Beyond single-word vocabulary*. Routledge.

Hirsh, D., & Coxhead, A. (2009). Ten ways of focussing on science-specific vocabulary in EAP classrooms. *English Australia Journal, 25*(1), 5–16.

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, *8*(2), 689–696.

Hirvela, A. (1997). "Disciplinary portfolios" and EAP writing instruction. *English for Specific Purposes*, *16*(2), 83–100. https://doi.org/10.1016/S0889-4906(96)00037-3

Holley, K. (2017). Interdisciplinary curriculum and learning in higher education. In *Oxford research encyclopedia of education*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190264093.013.138

Hsu, W. (2013). Bridging the vocabulary gap for EFL medical undergraduates: The establishment of a medical word list. *Language Teaching Research*, *17*(4), 454–484. https://doi.org/10.1177/1362168813494121

Hsu, W. (2014). Measuring the vocabulary load of engineering textbooks for EFL undergraduates. *English for Specific Purposes*, *33*(1), 54–65. https://doi.org/10.1016/j.esp.2013.07.001

Hsu, W. (2018). The most frequent BNC/COCA mid- and low-frequency word families in English-medium traditional Chinese medicine (TCM) textbooks. *English for Specific Purposes*, *51*, 98–110. https://doi.org/10.1016/j.esp.2018.04.001

Hu, H. C., & Nation, P. (2000). Unknown word density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403–430. https://nflrc.hawaii.edu/rfl/item/43

Huibregtse, I., Admiraal, W., & Meara, P. (2002). Scores on a yes-no vocabulary test: correction for guessing and response style. *Language Testing*, *19*(3), 227–245. https://doi.org/10.1191/0265532202lt229oa

Hunt, A., & Beglar, D. (2005). A framework for developing EFL reading vocabulary. *Reading in a Foreign Language*, *17*(1), 23–59.

Hyland, K. (2000). Hedges, boosters and lexical invisibility: Noticing modifiers in academic texts. *Language Awareness*, *9*(4), 179–197. https://doi.org/10.1080/09658410008667145

Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing*. University of Michigan Press. https://doi.org/10.3998/mpub.6719

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies, 7*(2), 173–192. https://doi.org/10.1177/146144560505036

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*(1), 4–21. https://doi.org/10.1016/j.esp.2007.06.001

Hyland, K. (2010). Researching writing. In L. Paltridge & A. Phakiti (Eds.), *Continuum companion to research methods in applied linguistics* (pp. 191–204). Bloomsbury Publishing.

Hyland, K. (2015). *Academic Written English*. Shanghai: Shanghai Foreign Language Education Press.

Hyland, K. (2016). Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing*, *31*, 58–69. https://doi.org/10.1016/j.jslw.2016.01.005

Hyland, K., & Tse, P. (2005). Hooking the reader: A corpus study of evaluative that in abstracts. *English for Specific Purposes*, *24*(2), 123–139. https://doi.org/10.1016/j.esp.2004.02.002

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly*, *41*(2), 235–253. https://doi.org/10.2307/40264352

Hyland, K., & Shaw, P. (Eds.). (2016).*The Routledge handbook of English for academic purposes*. Routledge. https://doi.org/10.4324/9781315657455

Jiang, F., & Hyland, K. (2015). "The fact that": Stance nouns in disciplinary writing. *Discourse Studies*, *17*(5), 529–550. https://doi.org/10.1177/1461445615590719

Jin, N. Y., Ling, L. Y., Tong, C. S., Sahiddan, N., Philip, A., Azmi, N. H. N., & Tarmizi, M. A. A. (2013). Development of the engineering technology word list for vocational schools in Malaysia. *International Education Research, 1*(1), 43–49.

Johns, T. (2002). Data-driven learning: The perpetual challenge. In B. Kettemann & G. Marko (eds.), *Teaching and learning by doing corpus analysis: Proceedings of the fourth international conference on teaching and language corpora* (pp. 105–117). Brill.

Jordan, G., & Gray, H. (2019). We need to talk about coursebooks. *ELT Journal*, *73*(4), 438–446. https://doi.org/10.1093/elt/ccz038

Juilland, A., & Chang-Rodrigues, C. (1964). Methods for measuring dispersion in linguistic data. *Language, 40*(1), 1–16.

Kachru, B. B. (1992). *The other tongue: English across cultures* (2nd ed.). University of Illinois Press.

Kahu, E. R., & Nelson, K. (2018). Student engagement in the educational interface: Understanding the mechanisms of student success. *Higher Education Research & Development, 37*(1), 58–71. https://doi.org/10.1080/07294360.2017.1344197

Kang, Y., & Cheng, X. (2014). Teacher learning in the workplace: A study of the relationship between a novice EFL teacher's classroom practices and cognition development. *Language Teaching Research*, *18*(2), 169–186. https://doi.org/10.1177/1362168813505939

Kennedy, G. D. (1998). *An introduction to corpus linguistics*. Longman.

Kennedy, G. (2003). Amplifier collocations in the British national corpus: Implications for English language teaching. *TESOL Quarterly*, *37*(3), 467–487. https://doi.org/10.2307/3588400

Khan, R. M. I., Radzuan, N. R. M., Shahbaz, M., Ibrahim, A. H., & Mustafa, G. (2018). The role of vocabulary knowledge in speaking development of Saudi EFL learners. *Arab World English Journal, 9*(1), 406–418. https://dx.doi.org/10.24093/awej/vol9no1.28

Knoch, U. (2014). Using subject specialists to validate an ESP rating scale: The case of the International Civil Aviation Organization (ICAO) rating scale. *English for Specific Purposes*, *33*(1), 77–86. https://doi.org/10.1016/j.esp.2013.08.002

Kress, G. (1989). *Linguistic processes in sociocultural practice*. Oxford, UK: Oxford University Press.

Kwary, D. A. (2011). A hybrid method for determining technical vocabulary. *System (Linköping)*, *39*(2), 175–185. https://doi.org/10.1016/j.system.2011.04.003

Kwary, D. A., Ratri, D., & Artha, A. F. (2017). Lexical bundles in journal articles across academic disciplines. *Indonesian Journal of Applied Linguistics*, *7*(1), 132–140. https://doi.org/10.17509/ijal.v7i1.6866

LaFlair, G. T., Egbert, J., Plonsky, L., (2015). A practical guide to bootstrapping descriptive statistics, correlations, T tests, and ANOVAs. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 46–77). Routledge. https://doi.org/10.4324/9781315870908-4

Lancaster, Z. (2016). Expressing stance in undergraduate writing: Discipline-specific and general qualities. *Journal of English for Academic Purposes*, *23*, 16–30. https://doi.org/10.1016/j.jeap.2016.05.006

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, *33*(2), 363–374. https://doi.org/10.2307/2529786

Laufer, B. (1987). A case for vocabulary in EAP reading comprehension materials. In Cornu, A.-M., Vanparijs, J. Delaheye M., & Baten L. (Eds.), *Beads or bracelet? How do we approach LSP. Selected papers from the 5th European symposium on language for specific purposes* (pp. 284–291). Oxford University Press.

Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). Multilingual Matters.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review*, *59*(4), 567–587. https://doi.org/10.3138/cmlr.59.4.567

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307–322. https://doi.org/10.1093/applin/16.3.307

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*(1), 36–55. https://doi.org/10.1191/026553299672614616

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*(1), 15–30. https://doi.org/10125/66648

Laufer, B., & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC Journal, 28*(1), 89–108. https://doi.org/10.1177/003368829702800106

Le, C. N. N., & Miller, J. (2025). Developing and piloting SemiMed—A resource for semi-technical medical vocabulary. *Research Methods in Applied Linguistics, 4*(3), 100239. https://doi.org/10.1016/j.rmal.2025.100239

Lee, J. J., & Casal, J. E. (2014). Metadiscourse in results and discussion chapters: A cross-linguistic analysis of English and Spanish thesis writers in engineering. *System (Linköping)*, *46*, 39–54. https://doi.org/10.1016/j.system.2014.07.009

Leeke, P., & Shaw, P. (2000). Learners' independent records of vocabulary. *System (Linköping)*, *28*(2), 271–289. https://doi.org/10.1016/S0346-251X(00)00012-9

Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, *22*, 42–53. https://doi.org/10.1016/j.jeap.2016.01.008

Lei, L., & Liu, D. (2018). The academic English collocation list: A corpus-driven study. *International Journal of Corpus Linguistics*, *23*(2), 216–243. https://doi.org/10.1075/ijcl.16135.lei

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, *18*(2), 85–102. https://doi.org/10.1016/j.jslw.2009.02.001

Liu, D., Lei, L., (2020). Technical vocabulary. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 111–124). Routledge. https://doi.org/10.4324/9780429291586-8

Liu, J., & Han, L. (2015). A corpus-based environmental academic word list building and its validity test. *English for Specific Purposes*, *39*, 1–11. https://doi.org/10.1016/j.esp.2015.03.001

Loghmani, Z., Ghonsooly, B., & Ghazanfari, M. (2020). Engagement in doctoral dissertation discussion sections written by English native speakers. *Journal of English for Academic Purposes*, *45*, 100851–13. https://doi.org/10.1016/j.jeap.2020.100851

Long, M. H. (Ed.). (2005). *Second language needs analysis*. Cambridge University Press.

Louber, I., & Troudi, S. (2019). "Most of the teaching is in Arabic anyway": English as a medium of instruction in Saudi Arabia, between de facto and official language policy. *International Journal of Bias, Identity and Diversities in Education, 4*(2), 59–73. https://doi.org/10.4018/IJBIDE.2019070105

Lu, C. (2018). *Investigating knowledge and use of technical vocabulary in traditional Chinese medicine* [Doctoral dissertation]. Te Herenga Waka- –Victoria University of Wellington. https://openaccess.wgtn.ac.nz/articles/thesis/Investigating_knowledge_and_use_of_technical_vocabulary_in_Traditional_Chinese_Medicine/17131862?file=31677065

Lu, C., & Coxhead, A. (2020). Vocabulary in traditional Chinese medicine: Insights from corpora. *ITL – International Journal of Applied Linguistics*, *171*(1), 34–61. https://doi.org/10.1075/itl.18020.lu

Lu, C., & Dang, T. (2022). Vocabulary in EAP learning materials: What can we learn from teachers, learners, and corpora? *System, 106*, 102791. https://doi.org/10.1016/j.system.2022.102791

Lu, C., & Durrant, P. (2017). A corpus-based lexical analysis of Chinese medicine research articles. *The Asian Journal of Applied Linguistics, 4*(1), 3–15. https://caes.hku.hk/ajal/index.php/ajal/article/view/437

Lynn, R. W. (1973). Preparing word-lists: A suggested method. *RELC Journal,4*(1), 25–28. https://doi.org/10.1177/003368827300400103

Macaro, E., Tian, L., & Chu, L. (2020). First and second language use in English medium instruction contexts. *Language Teaching Research, 24*(3), 382–402. https://doi.org/10.1177/1362168818783231

Mackey, W. F. (1965). *Language teaching analysis*. Longmans.

Mahibur Rahman, M. (2011). English language teaching (ELT) in Saudi Arabia: A study of learners' needs analysis with special reference to community college, Najran University. *Language in India, 11*(4), 367–461.

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, *33*(3), 299–320. https://doi.org/10.1093/applin/ams010

Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System (Linköping)*, *52*, 1–13. https://doi.org/10.1016/j.system.2015.04.015

McCarthy, M. J., McCarten, J., & Sandiford, H. (2005a). *Touchstone 1 A workbook A level 1*. Cambridge University Press.

McCarthy, M. J., McCarten, J., & Sandiford, H. (2005b). *Touchstone 2 A workbook A Level 2*. Cambridge University Press.

McCarthy, M. J., McCarten, J., & Sandiford, H. (2005c). *Touchstone 3 A workbook A Level 3*. Cambridge University Press.

McEnery, T., & Brookes, G. (2024). Corpus linguistics and the social sciences. *Corpus Linguistics and Linguistic Theory*, *20*(3), 591–613. https://doi.org/10.1515/cllt-2024-0036

McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, Theory and Practice*. Cambridge University Press. https://doi.org/10.1017/CBO9780511981395

McGraw-Hill. (2003). *McGraw-Hill's dictionary of engineering*. McGraw-Hill.

Meara, P. (1989). Matrix models of vocabulary acquisition. *AILA Review*, 6, 66–74.

Meara, P. M. (1992). EFL vocabulary tests (2nd ed.). *Centre for Applied Language Studies.* https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=d8306ac3f4f3f2fb3bdac0b9c2c5662f6583a642

Meara, P. (1996a). *The classical research in vocabulary acquisition: Words, words, words*. Multilingual Matters. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=126fe3a92d35ea8e46201e30d0f51687e9e6e21e

Meara, P. (1996b). The vocabulary knowledge framework. Vocabulary Acquisition Research Group Virtual Library. https://www.lognostics.co.uk/vlibrary/meara1996c.pdf

Meara, P., & Bell, H. (2001). P-Lex: A simple and effective way of describing the lexical characteristics of short L2 tests. *Prospect, 16*(3), 5–19. https://eric.ed.gov/?id=EJ637430

Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing, 4*(2), 142–154. https://journals.sagepub.com/doi/abs/10.1177/026553228700400202

Merriam-Webster. (n.d.). *Merriam-Webster online dictionary*. Retrieved October 2023, from https://www.merriam-webster.com

Messick, S. (1989). *Educational measurement*. American Council on Education.

Mochida, K. & Harrington, M. (2006). The yes/no test as a measure of receptive vocabulary knowledge. *Language Testing*, *23*(1), 73–98. https://doi.org/10.1191/0265532206lt321oa

Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes*, *25*(2), 235–256. https://doi.org/10.1016/j.esp.2005.05.002

Muhammad, I., & Abdul Raof, A. H. (2020). Assessing the ESP needs of Saudi engineering undergraduates: A diagnostic view. *Arab World English Journal, 10,* 190–204. https://doi.org/10.24093/awej/vol10no3.13

Murphy, M. L. (2016). The Oxford handbook of the word (Book review). *International Journal of Lexicography*, *30*(4), 525–528. https://doi.org/10.1093/ijl/ecw013

Nasseri, M., & Thompson, P. (2021). Lexical density and diversity in dissertation abstracts: Revisiting English L1 vs. L2 text differences. *Assessing Writing*, *47*, 100511. https://doi.org/10.1016/j.asw.2020.100511

Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines, 5*(1), 12–25.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Heinle & Heinle.

Nation, P. (2000). Learning vocabulary in lexical sets: Dangers and guidelines. *TESOL Journal*, *9*(2), 6–10. https://doi.org/10.1002/j.1949-3533.2000.tb00239.x

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*(1), 59–82. https://doi.org/10.3138/cmlr.63.1.59

Nation, I. S. P. (2012). The BNC/COCA word family lists. Available at: http://www.victoria.ac.nz/lals/about/staff/paul-nation

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. John Benjamins Publishing.

Nation, I. S. P., Coxhead, A., Chung, T. M., & Quero, M. (2016). Specialised word lists. In I. S. P. Nation (Ed.), *Making and using word lists for language learning and testing* (pp. 145-151). John Benjamins Publishing.

Nation, P., & Kyongho, H. (1995). Where would general service vocabulary stop and special purposes vocabulary begin? *System (Linköping)*, *23*(1), 35–41. https://doi.org/10.1016/0346-251X(94)00050-G

Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6–19). Cambridge University Press.

Nation, I. S. P., & Webb, S. (2011). Content-based instruction and vocabulary learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 631–644). Routledge. https://doi.org/10.4324/9780203836507-43

Neuendorf, K. A. (2017). *The content analysis guidebook* (2nd ed.). SAGE Publications. https://doi.org/10.4135/9781071873045

Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, *21*(3), 298–320. https://doi.org/10.1177/1362168816639619

Nunnally, J., & Bernstein, I. (1994*). Psychometric theory* (3rd ed.). McGraw-Hill.

Nurmukhamedov, U., & Webb, S. (2019). Lexical coverage and profiling. *Language Teaching*, *52*(2), 188–200. https://doi.org/10.1017/S0261444819000028

Nesi, H. (2013). ESP and corpus studies. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 407–426). Wiley-Blackwell.

Oh, J., Lee, J., Lee, K., & Choi, K. (2000). Japanese term extraction using dictionary hierarchy and a machine translation system. *Terminology, 6*, 287–311.

Omidian, T., & Siyanova-Chanturia, A. (2020). Semantic prosody revisited. *TESOL Quarterly*, *54*(2), 512–524. https://doi.org/10.1002/tesq.557

Omidian, T., & Siyanova-Chanturia, A. (2021). Parameters of variation in the use of words in empirical research writing. *English for Specific Purposes*, *62*, 15–29. https://doi.org/10.1016/j.esp.2020.11.001

Paltridge, B., & Starfield, S. (2020). Change and continuity in thesis and dissertation writing: The evolution of an academic genre. *Journal of English for Academic Purposes*, *48*, 100910. https://doi.org/10.1016/j.jeap.2020.100910

Pellicer Sanchez, A. (2020). Expanding English vocabulary knowledge through reading: Insights from eye-tracking studies. *RELC Journal, 51*(1), 134–146. https://doi.org/10.1177/0033688220906904

Pellicer-Sánchez, A., & Schmitt, N. (2012). Scoring yes–no vocabulary tests: Reaction time vs. nonword approaches. *Language Testing*, *29*(4), 489–509. https://doi.org/10.1177/0265532212438053

Plonsky, L. (2015). *Advancing quantitative methods in second language research*. Routledge. https://doi.org/10.4324/9781315870908

Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Praninskas, J. (1972). *American university word list*. Longman.

Polgar, S., & Thomas, S. A. (2000). *Introduction to research in the health sciences* (4th ed.). Churchill Livingstone.

Qi, H. (2016). *A corpus-based comparison between the Academic Word List and the Academic Vocabulary List* [Master's thesis, The University of Western Ontario]. https://ir.lib.uwo.ca/etd/3721

Quero, B. (2015). *Estimating the vocabulary size of L1 Spanish ESP learners and the vocabulary load of medical textbooks* [Doctoral dissertation]. Te Herenga Waka –Victoria University of Wellington.

Quero, B., & Coxhead, A. (2018). Using a corpus-based approach to select medical vocabulary for an ESP course: The case for high-frequency vocabulary. In Y. Kirkgöz & K. Dikilitaş (Eds.), *Key issues in English for specific purposes in higher education* (pp. 51–75). https://link.springer.com/chapter/10.1007/978-3-319-70214-8_4

Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal, 19*, 12–25. https://journals.sagepub.com/doi/abs/10.1177/003368828801900202

Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.

Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly, 10*(1), 77–89. https://doi.org/10.2307/3585941

Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, *31*, 205–225. https://doi.org/10.1017/S0267190511000055

Sakata, N. (2019). Profiling vocabulary for proficiency development: Effects of input and general frequencies on L2 learning. *System (Linköping)*, *87*, 102167. https://doi.org/10.1016/j.system.2019.102167

Salager, F. (1983). The lexis of fundamental medical English: classificatory framework and rhetorical function (a statistical approach). *Reading in a Foreign Language, 1*, 54–64.

Salmons, J. (2014). *New social media, new social science ... and new ethical issues*. Available from https://www.slideshare.net/slideshow/nsmnss-report-on-research-ethics-final/44392946

Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning*, *48*(2), 281–317. https://doi.org/10.1111/1467-9922.00042

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan. https://doi.org/10.1057/9780230293977

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, *95*(1), 26–43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, *47*(4), 484–503. https://doi.org/10.1017/S0261444812000018

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55–88. https://doi.org/10.1177/026553220101800103

Scott, M. (1997). PC analysis of key words – and key key words. *System, 25*(2), 233–245. https://doi.org/10.1016/S0346-251X(97)00011-0

Scott, M. (1998). *WordSmith Tools manual* (Version 3.0). Oxford University Press.

Scott, M. (2012). *WordSmith Tools* (Version 6) [Computer software]. Lexical Analysis Software. https://www.lexically.net/wordsmith/

Shin, D., & Nation, P. (2008). Beyond single words: The most frequent collocations in spoken English. *ELT Journal*, *62*(4), 339–348. https://doi.org/10.1093/elt/ccm091

Siegel, J. (2020). Comprehension in English medium instruction (EMI) lectures: On the impact of lecturer L2 English use. *Language Learning in Higher Education, 10*(1), 73–91. https://doi.org/10.1515/cercles-2020-2005

Simbuka, S., Hamied, F. A., Sundayana, W., & Kwary, D. A. (2019). A corpus-based study on the technical vocabulary of Islamic religious studies. *TEFLIN Journal*, *30*(1), 47–71. https://doi.org/10.15639/teflinjournal.v30i1/47-71

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, *31*(4), 487–512. https://doi.org/10.1093/applin/amp058

Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, *36*(2), 139–152. https://doi.org/10.1080/09571730802389975

Stamatović, M. V., Bratić, V., & Lakić, I. (2020). Vocabulary of L1 and L2 graduation theses written by English philology students: Academic writing of Montenegrin and US students compared. *ELOPE (Tiskana Izd.)*, *17*(2), 101–113. https://doi.org/10.4312/elope.17.2.101-113

Staples, S., & Reppen, R. (2016). Understanding first-year L2 writing: A lexico-grammatical analysis across L1s, genres, and language ratings. *Journal of Second Language Writing*, *32*, 17–35. https://doi.org/10.1016/j.jslw.2016.02.002

Sun, S. A., & Crosthwaite, P. (2022). “The findings might not be generalizable”: Investigating negation in the limitations sections of PhD theses across disciplines. *Journal of English for Academic Purposes*, *59*, 101155. https://doi.org/10.1016/j.jeap.2022.101155

Sun, Y., & Dang, T. N. Y. (2020). Vocabulary in high-school EFL textbooks: Texts and learner knowledge. *System (Linköping)*, *93*, 102279. https://doi.org/10.1016/j.system.2020.102279

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Szudarski, P. (2018). *Corpus linguistics for vocabulary: A guide for research*. Routledge.

Tamran, R. (2016). *Teaching English as a second language in Saudi Arabia* [Unpublished project paper]. Effat University, Jeddah.

Tardy, C. (2009). *Building genre knowledge.* Parlor Press.

Taymaz, N. (2021). A corpus-based comparison of use of hedges and boosters by Turkish ELT MA and PhD students*. Journal of Language and Linguistic Studies, 17*, 33–49.

Therova, D. (2020). Review of academic word lists. *TESL-EJ, 24*(1).

Therova, D. (2021). *A corpus-based study of academic vocabulary in foundation-level students' assessed academic writing at a UK university.* Open University (United Kingdom).

Thompson, J. A. (2005). Proactive personality and job performance: A social capital perspective. *Journal of Applied Psychology*, *90*(5), 1011–1017. https://doi.org/10.1037/0021-9010.90.5.1011

Thompson, P. (1999). Exploring the contexts of writing: Interviews with PhD supervisors. In P. Thompson (Ed.), *Issues in EAP writing research and instruction* (pp. 37–54). CALS.

Thompson, P. (2016). Genre approaches to theses and dissertations. In K. Hyland (Ed.), *The Routledge handbook of English for academic purposes* (pp. 379–391). Routledge. https://doi.org/10.4324/9781315657455

Thompson, P., & Hunston, S. (2019). *Interdisciplinary research discourse: Corpus investigations into environment journals*. Routledge. https://doi.org/10.4324/9781003009344

Tidball, F., & Treffers-Daller, J. (2008). Analysing lexical richness in French learner language: What frequency lists and teacher judgements can tell us about basic and advanced words. *Journal of French Language Studies*, *18*(3), 299–313. https://doi.org/10.1017/S0959269508003463

Tongpoon-Patanasorn, A. (2018). Developing a frequent technical words list for finance: A hybrid approach. *English for Specific Purposes*, *51*, 45–54. https://doi.org/10.1016/j.esp.2018.03.002

Torr, J., & Scott, C. (2006). Learning 'special words' technical vocabulary in the talk of adults and preschoolers during shared reading. *Journal of Early Childhood Research, 4*(2), 153–167. https://doi.org/10.1177/1476718X06063534

Trang, N. H., Truong, D. V., & Ha, H. T. (2023). Quantifying vocabulary learning belief and strategy – A validation study of the Vietnamese version of Gu's (2018) vocabulary learning questionnaire. *Heliyon*, *9*(5), e16009–e16009. https://doi.org/10.1016/j.heliyon.2023.e16009

ur Rahman, M. M., & Alhaisoni, E. (2013). Teaching English in Saudi Arabia: Prospects and challenges. *Academic Research International, 4*(1), 112–118.

Valipouri, L., & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes*, *12*(4), 248–263. https://doi.org/10.1016/j.jeap.2013.07.001

Van Lancker-Sidtis, D., & Rallon, G. (2004). Tracking the incidence of formulaic expressions in everyday speech: Methods for classification and verification. *Language & Communication*, *24*(3), 207–240. https://doi.org/10.1016/j.langcom.2004.02.003

Van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, *34*(4), 457–479. https://doi.org/10.1093/applin/ams074

Vilkaitė-Lozdienė, L., Schmitt, N., & Webb, S. (2020). Frequency as a guide for vocabulary usefulness: High-, mid-, and low-frequency words. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 81–96). Routledge. https://doi.org/10.4324/9780429291586-6

Wang, J., Liang, S., & Ge, G. (2008). Establishment of a medical academic word list. *English for Specific Purposes*, *27*(4), 442–458. https://doi.org/10.1016/j.esp.2008.05.003

Ward, J. (1999). How large a vocabulary do EAP engineering students need? *Reading in a Foreign Language*, *12*(2), 309–323. https://doi.org/10125/66967

Ward, J. (2007). Collocation and technicality in EAP engineering. *Journal of English for Academic Purposes*, *6*(1), 18–35. https://doi.org/10.1016/j.jeap.2006.10.001

Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes*, *28*(3), 170–182. https://doi.org/10.1016/j.esp.2009.04.001

Watson Todd, R. (2017). An opaque engineering word list: Which words should a teacher focus on? *English for Specific Purposes*, *45*, 31–39. https://doi.org/10.1016/j.esp.2016.08.003

Webb, S. (2017). *How vocabulary is learned*. Oxford University Press.

Webb, S. A., & Chang, A. C.-S. (2012). Second language vocabulary growth. *RELC Journal*, *43*(1), 113–126. https://doi.org/10.1177/0033688212439367

Webb, S., & Nation, P. (2008). Evaluating the vocabulary load of written text. *TESOLANZ Journal*, *16*, 1–9.

Webb, S. A., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press.

Webb, S., & Paribakht, T. S. (2015). What is the relationship between the lexical profile of test items and performance on a standardized English proficiency test? *English for Specific Purposes*, *38*, 34–43. https://doi.org/10.1016/j.esp.2014.11.001

Webb, S., & Rodgers, M. P. H. (2009a). The lexical coverage of movies. *Applied Linguistics*, *30*(3), 407–427. https://doi.org/10.1093/applin/amp010

Webb, S., & Rodgers, M. P. H. (2009b). Vocabulary demands of television programs. *Language Learning*, *59*(2), 335–366. https://doi.org/10.1111/j.1467-9922.2009.00509.x

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL – International Journal of Applied Linguistics*, *168*(1), 33–69. https://doi.org/10.1075/itl.168.1.02web

West, M. (1953). *A general service list of English words*. Longman, Green and Co.

Wilkins, D. A. (1972). *Linguistics in language teaching*. Edward Arnold.

Williams, R. (1985). Teaching vocabulary recognition strategies in ESP reading. *ESP Journal, 4*, 121–131.

Willis, M., & Ohashi, Y. (2012). A model of L2 vocabulary learning and retention. *Language Learning Journal*, *40*(1), 125–137. https://doi.org/10.1080/09571736.2012.658232

Wood, D. C., & Appel, R. (2014). Multiword constructions in first year business and engineering university textbooks and EAP textbooks. *Journal of English for Academic Purposes*, *15*, 1–13. https://doi.org/10.1016/j.jeap.2014.03.002

Wood, D. (2020). Classifying and identifying formulaic language. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 30–45). Routledge. https://doi.org/10.4324/9780429291586-3

Woodward-Kron, R. (2008). More than just jargon – The nature and role of specialist language in learning disciplinary knowledge. *Journal of English for Academic Purposes*, *7*(4), 234–249. https://doi.org/10.1016/j.jeap.2008.10.004

Wray, A. (2000). Formulaic sequences in second language teaching: principle and practice. *Applied Linguistics*, *21*(4), 463–489. https://doi.org/10.1093/applin/21.4.463

Wray, A. (2002). Formulaic language in computer-supported communication: Theory meets reality. *Language Awareness*, *11*(2), 114–131. https://doi.org/10.1080/09658410208667050

Xing, P., & Fulcher, G. (2007). Reliability assessment for two versions of Vocabulary Levels Tests. *System (Linköping)*, *35*(2), 182–191. https://doi.org/10.1016/j.system.2006.12.009

Xiao, W., & Sun, S. (2020). Dynamic lexical features of PhD theses across disciplines: A text mining approach. *Journal of Quantitative Linguistics*, *27*(2), 114–133. https://doi.org/10.1080/09296174.2018.1531618

Xue, G., & Nation, P. (1984). A university word list. *Language Learning and Communication, 3*(2), 215–229.

Yang, M.-N. (2015). A nursing academic word list. *English for Specific Purposes*, *37*(1), 27–38. https://doi.org/10.1016/j.esp.2014.05.003

Appendix A: **Teachers' Rating for Potential Engineering Technical Vocabulary List and Sematic Scale**

#	Keyword	Rater 1				Rater 2				Rater 3				T V	NO T	TV. 1
		1	2	3	4	1	2	3	4	1	2	3	4			
1	ACADEMIC	✓				✓				✓				0	3	
2	ACCIDENTS	✓				✓					✓			1	2	
3	ACCREDITED		✓			✓					✓			2	1	✓
4	ACCURATE		✓				✓				✓			3	0	✓
5	ACHIEVE	✓				✓					✓			1	2	
6	ACHIEVED	✓				✓					✓			1	2	
7	ACHIEVING	✓				✓					✓			1	2	
8	ACTIVATION		✓					✓				✓		3	0	✓
9	ACTUAL	✓				✓					✓			1	2	
10	ADDITION	✓				✓					✓			1	2	
11	ADDITIONALLY	✓				✓					✓			1	2	
12	ADOPTION	✓					✓				✓			2	1	✓
13	ADVANTAGES	✓				✓				✓				0	3	
14	AFFECT		✓			✓					✓			2	1	✓
15	AFFECTING		✓			✓					✓			2	1	✓
16	AFFECTS		✓			✓					✓			2	1	✓
17	AGENTS			✓			✓					✓		3	0	✓
18	AGGREGATE			✓					✓			✓		3	0	✓
19	AIRLINE		✓				✓			✓				2	1	✓
20	AIRLINES		✓				✓			✓				2	1	✓
21	ALGORITHMS				✓	✓							✓	2	1	✓
22	ALLOWS	✓				✓				✓				0	3	
23	ALLOYS				✓				✓				✓	3	0	
24	ALTERNATIVES		✓			✓					✓			2	1	✓
25	ALUMINUM				✓			✓			✓			3	0	✓
26	ANALYZED		✓			✓					✓			2	1	✓
27	ANALYZING		✓			✓					✓			2	1	✓
28	ANGLES			✓				✓				✓		2	1	✓
29	ANNULUS				✓				✓				✓	3	0	✓
30	ANTENNA				✓				✓				✓	3	0	✓
31	ANTENNAS				✓				✓				✓	3	0	✓
32	APPENDIX			✓		✓				✓				1	2	✓
33	APPLICABLE			✓		✓					✓			2	1	
34	APPLICATION			✓				✓			✓			3	0	✓
35	APPLICATIONS			✓				✓			✓			3	0	✓
36	APPLYING		✓			✓					✓			2	1	✓
37	APPROACHES			✓			✓			✓				2	1	✓
38	APPROXIMATELY		✓				✓				✓			3	0	✓
39	ARMCHAIR							✓		✓				3	0	✓
40	ARRAYS				✓			✓					✓	3	0	✓
41	ASSIGNED	✓				✓				✓				0	3	✓
42	AUTHORS	✓				✓				✓				0	3	
43	BACHELOR'S	✓				✓				✓				0	3	
44	BANDS			✓				✓					✓	3	0	✓
45	BARRIERS			✓				✓				✓		3	0	✓

46	BASED			✓		✓				✓				1	2	
47	BATTERIES		✓					✓				✓		3	0	✓
48	BEAMS				✓				✓				✓	3	0	✓
49	BEHAVIOR			✓		✓				✓				1	2	
50	BRINE			✓					✓			✓		3	0	✓
51	BUSES	✓						✓		✓				1	2	
52	CALCULATE	✓				✓					✓			1	2	
53	CALCULATED	✓				✓					✓			1	2	
54	CALCULATING	✓				✓					✓			1	2	
55	CALCULATION	✓					✓				✓			2	1	✓
56	CALCULATIONS	✓					✓				✓			2	1	✓
57	CAMPUS	✓				✓					✓			1	2	
58	CANCER	✓				✓				✓				0	3	
59	CAPABILITIES			✓			✓			✓				2	1	✓
60	CAPABILITY			✓				✓		✓				2	1	✓
61	CAPTURE		✓			✓				✓				1	2	
62	CARRIERS			✓			✓					✓		3	0	✓
63	CATALYSTS				✓			✓					✓	3	0	✓
64	CATEGORIES		✓			✓					✓			2	1	✓
65	CATEGORY		✓			✓					✓			2	1	✓
66	CAUSES		✓			✓				✓				1	2	
67	CHALLENGES			✓		✓				✓				1	2	
68	CHANNELS			✓				✓				✓		3	0	✓
69	CHAPTER	✓				✓				✓				0	3	
70	CHARACTERISTICS	✓				✓				✓				0	3	
71	CHARACTERIZED	✓				✓				✓				0	3	
72	CLASSIFIED			✓		✓				✓				1	2	
73	CODES			✓				✓			✓			3	0	✓
74	COLLECTED		✓			✓				✓				1	2	
75	COMBINATIONS		✓					✓		✓				2	1	✓
76	COMBINING		✓					✓		✓				2	1	✓
77	COMMERCE	✓				✓				✓				0	3	
78	COMMONLY	✓				✓				✓				0	3	
79	COMMUNICATION	✓				✓				✓				0	3	
80	COMPACT		✓					✓			✓			3	0	✓
81	COMPARE	✓				✓				✓				0	3	
82	COMPARED	✓				✓				✓				0	3	
83	COMPARING	✓				✓				✓				0	3	
84	COMPLEXITY	✓						✓		✓				1	2	
85	COMPONENTS		✓					✓				✓		3	0	✓
86	COMPOSITES		✓					✓			✓			3	0	✓
87	COMPUTATIONAL		✓					✓		✓				2	1	✓
88	CONCLUDED	✓				✓				✓				0	3	
89	CONCLUSION	✓				✓				✓				0	3	
90	CONDITIONS			✓			✓			✓				2	1	✓
91	CONDUCTED		✓			✓				✓				1	2	
92	CONDUCTING		✓			✓				✓				1	2	
93	CONFIGURATIONS				✓			✓		✓				2	1	✓
94	CONNECT		✓					✓		✓				2	1	✓
95	CONNECTED		✓					✓		✓				2	1	✓
96	CONSIDERED	✓				✓				✓				0	3	

97	CONSIST		✓			✓				✓				0	3	
98	CONSISTS		✓			✓				✓				1	2	
99	CONSTRUCT		✓					✓				✓		3	0	✓
100	CONTAINS		✓			✓				✓				0	3	
101	CONTINUOUSLY	✓				✓				✓				0	3	
102	CONVERGENCE			✓		✓				✓				1	2	
103	CONVERT			✓				✓		✓				2	1	✓
104	COOLANT				✓				✓				✓	3	0	✓
105	CORRELATIONS			✓		✓					✓			2	1	✓
106	CORRESPONDING		✓			✓				✓				1	2	
107	COUNTS		✓			✓				✓				1	2	
108	CRACKS			✓				✓				✓		3	0	✓
109	CUSTOMER	✓				✓				✓				0	3	
110	CUSTOMERS	✓				✓				✓				0	3	
111	CYCLES			✓				✓		✓				2	1	✓
112	DAMAGE	✓				✓				✓				0	3	
113	DECREASE		✓			✓				✓				1	2	
114	DECREASED		✓			✓				✓				1	2	
115	DECREASES		✓			✓				✓				1	2	
116	DECREASING		✓			✓				✓				1	2	
117	DEFECT		✓					✓		✓				2	1	✓
118	DEFECTS		✓					✓		✓				2	1	✓
119	DEFINE	✓				✓				✓				0	3	
120	DEFINED	✓				✓				✓				0	3	
121	DEFINES	✓				✓				✓				0	3	
122	DEFINING	✓				✓				✓				0	3	
123	DEFINITIONS	✓				✓				✓				0	3	
124	DELAYS		✓				✓			✓				2	1	✓
125	DEPENDENT			✓		✓				✓				1	2	
126	DEPENDING	✓				✓				✓				0	3	

127	DEPENDS	✓				✓				✓				0	3	
128	DEPICTED	✓				✓				✓				0	3	
129	DEPOSITION		✓					✓					✓	3	0	✓
130	DESIGNED			✓		✓				✓				1	2	
131	DESIGNING			✓			✓			✓				2	1	✓
132	DESTINATION		✓			✓				✓				1	2	
133	DETECT		✓			✓				✓				1	2	
134	DETECTED		✓			✓				✓				1	2	
135	DETERMINE	✓				✓				✓				0	3	
136	DETERMINING	✓				✓				✓				0	3	
137	DEVELOPED	✓				✓				✓				0	3	
138	DEVICES	✓						✓		✓				1	2	
139	DIAGNOSTIC		✓			✓				✓				1	2	
140	DIALYSIS		✓			✓							✓	2	1	✓
141	DIAMETERS		✓					✓		✓				2	1	✓
142	DIFFERENT	✓				✓				✓				0	3	
143	DIMENSIONAL		✓					✓		✓				2	1	✓
144	DIMENSIONS		✓					✓		✓				2	1	✓
145	DISADVANTAGES	✓				✓				✓				0	3	
146	DISAGREE	✓				✓				✓				0	3	
147	DISCUSSED	✓				✓				✓				0	3	
148	DOMAINS			✓				✓		✓				2	1	✓
149	DOSE			✓			✓				✓			3	0	✓
150	EFFICIENT		✓			✓				✓				1	2	
151	ELECTRICITY		✓					✓		✓				2	1	✓
152	ELEMENTS		✓					✓		✓				2	1	✓
153	ELIMINATE	✓				✓						✓		1	2	
154	EMISSION		✓				✓						✓	3	0	✓
155	EMISSIONS		✓				✓						✓	3	0	✓

156	EMPLOYEE	✓				✓				✓				0	3	
157	EMPLOYEES	✓				✓				✓				0	3	
158	ENERGIES		✓					✓				✓		3	0	✓
159	ENHANCE		✓			✓				✓				1	2	
160	ENHANCED		✓			✓				✓				1	2	
161	ENTERPRISES		✓					✓		✓				2	1	✓
162	ENTITIES		✓					✓			✓			3	0	✓
163	ENTITY		✓					✓			✓			3	0	✓
164	ENTREPRENEURS		✓			✓				✓				1	2	
165	ENTREPRENEURSHIP		✓			✓				✓				1	2	
166	ENVIRONMENTS	✓						✓		✓				1	2	
167	EPITHERMAL				✓				✓				✓	3	0	✓
168	EQUALIZATION				✓			✓					✓	3	0	✓
169	ERGONOMIC				✓				✓	✓				2	1	✓
170	EROSION				✓				✓				✓	3	0	✓
171	ERRORS		✓					✓		✓				2	1	✓
172	ESTIMATE		✓					✓		✓				2	1	✓
173	ESTIMATION		✓					✓		✓				2	1	✓
174	EVALUATE		✓			✓				✓				1	2	
175	EVALUATED		✓			✓				✓				1	2	
176	EVALUATING		✓			✓				✓				1	2	
177	EVALUATION		✓					✓		✓				2	1	✓
178	EXCEED		✓					✓		✓				2	1	✓
179	EXECUTIVES		✓					✓		✓				2	1	✓
180	EXPERIMENTAL		✓				✓				✓			3	0	✓
181	EXPERIMENTALLY		✓				✓				✓			3	0	✓
182	EXPERIMENTS		✓				✓				✓			3	0	✓
183	EXTRACTED		✓			✓					✓			2	1	✓
184	FABRICATED		✓				✓						✓	3	0	✓

185	FACILITY	✓						✓		✓				1	2	
186	FACTORS	✓				✓				✓				0	3	
187	FADING		✓					✓					✓	3	0	✓
188	FAILURES		✓					✓		✓				2	1	✓
189	FAULTS		✓					✓		✓				2	1	✓
190	FEATURE		✓					✓		✓				2	1	✓
191	FEATURES	✓						✓		✓				1	2	
192	FIBERS		✓			✓					✓			2	1	✓
193	FILTERS		✓					✓			✓			3	0	✓
194	FINS			✓				✓				✓		3	0	✓
195	FIXED		✓			✓				✓				2	1	✓
196	FLIGHTS	✓					✓			✓				1	2	
197	FLOWS			✓		✓				✓				1	2	
198	FOCUSED		✓			✓				✓				1	2	
199	FOCUSES		✓			✓				✓				1	2	
200	FOLLOWS	✓				✓				✓				0	3	
201	FORECASTING		✓				✓			✓				2	1	✓
202	FORECASTS		✓				✓			✓				2	1	✓
203	FORMULATION			✓				✓		✓				2	1	✓
204	FOSSIL		✓					✓		✓			✓	3	0	✓
205	FREQUENCIES			✓				✓					✓	3	0	✓
206	FUELS		✓					✓		✓				2	1	✓
207	FUNCTIONS		✓				✓			✓				2	1	✓
208	FURTHERMORE	✓				✓				✓				0	3	
209	GASES	✓						✓			✓			2	1	✓
210	GENERATE		✓			✓							✓	2	1	✓
211	GENERATED		✓			✓							✓	2	1	✓
212	GENERATORS				✓			✓			✓			3	0	✓
213	GEOMETRY				✓			✓					✓	3	0	✓

214	GLOBAL	✓					✓			✓				1	2	
215	GRADIENT				✓			✓					✓	3	0	✓
216	GRADUATE	✓				✓				✓				0	3	
217	HARMONICS				✓			✓					✓	3	0	✓
218	HAZARDS		✓					✓			✓			3	0	✓
219	HEALTHCARE	✓				✓				✓				0	3	
220	HENCE	✓				✓				✓				0	3	
221	HIGH	✓				✓				✓				0	3	
222	HIGHER	✓				✓				✓				0	3	
223	HIGHEST	✓				✓				✓				0	3	
224	HOLLOW		✓					✓				✓		3	0	✓
225	HOSPITALS	✓				✓				✓				0	3	
226	HYDRO				✓			✓					✓	3	0	✓
227	HYDROGEN				✓				✓		✓			3	0	✓
228	IDENTIFY	✓				✓				✓				0	3	
229	IDENTIFYING	✓				✓				✓				0	3	
230	ILLUSTRATED	✓				✓				✓				0	3	
231	ILLUSTRATES	✓				✓				✓				0	3	
232	IMAGERY	✓							✓	✓				1	2	
233	IMAGES			✓				✓		✓				2	1	✓
234	IMPACTS	✓						✓		✓				1	2	
235	IMPLEMENT	✓				✓				✓				0	3	
236	IMPLEMENTED	✓				✓				✓				0	3	
237	IMPLEMENTING	✓				✓				✓				0	3	
238	IMPROVE	✓				✓				✓				0	3	
239	IMPROVED	✓				✓				✓				0	3	
240	IMPROVEMENT	✓					✓			✓				1	2	
241	IMPROVING	✓				✓				✓				0	3	
242	INCIDENT	✓				✓				✓				0	3	

243	INCREASE		✓			✓				✓				1	2	
244	INCREASES		✓			✓				✓				1	2	
245	INCREASING		✓			✓				✓				1	2	
246	INCUBATORS				✓				✓			✓		3	0	✓
247	INDICATES	✓				✓				✓				0	3	
248	INDICATORS		✓				✓				✓			3	0	✓
249	INFLUENCES	✓				✓				✓				0	3	
250	INFRASTRUCTURE				✓			✓			✓			3	0	✓
251	INNOVATION		✓				✓			✓				2	1	✓
252	INPUTS			✓				✓		✓				2	1	✓
253	INSTALLED		✓				✓						✓	3	0	✓
254	INTERRUPTION		✓					✓		✓				2	1	✓
255	INTERVAL		✓				✓					✓		3	0	✓
256	INTRODUCTION	✓				✓				✓				0	3	
257	INVESTIGATE	✓				✓				✓				0	3	
258	INVESTIGATED	✓				✓				✓				0	3	
259	IRRADIANCE				✓				✓	✓				2	1	✓
260	ITEM	✓						✓		✓				1	2	
261	ITEMS	✓						✓		✓				1	2	
262	ITERATION			✓					✓				✓	3	0	✓
263	ITERATIONS			✓					✓				✓	3	0	✓
264	KERNEL				✓			✓					✓	3	0	✓
265	LAMINAR				✓				✓			✓		3	0	✓
266	LAYERS			✓				✓				✓		3	0	✓
267	LEADS	✓						✓					✓	2	1	✓
268	LICENSE	✓						✓		✓				1	2	
269	LIMITATIONS	✓					✓			✓				1	2	
270	LITERATURE	✓				✓				✓				0	3	
271	LOADS			✓				✓				✓		3	0	✓

272	LOCALIZATION		✓					✓			✓			3	0	✓
273	LOCATED		✓			✓				✓				1	2	
274	LOCATIONS		✓				✓			✓				2	1	✓
275	LOOPS				✓				✓			✓		3	0	✓
276	LOSSES	✓				✓				✓				0	3	
277	LOWEST	✓				✓				✓				0	3	
278	MAGNITUDE			✓				✓				✓		3	0	✓
279	MANUFACTURING			✓				✓		✓				2	1	✓
280	MAPPING			✓				✓		✓				2	1	✓
281	MASSES				✓			✓				✓		3	0	✓
282	MATCHING			✓				✓		✓				2	1	✓
283	MATERIALS		✓					✓		✓				2	1	✓
284	MATHEMATICAL		✓			✓					✓			2	1	✓
285	MAXIMIZE	✓				✓				✓				0	3	
286	MEASURE	✓					✓			✓				1	2	
287	MEASUREMENTS	✓					✓			✓				1	2	
288	MECHANISM		✓					✓			✓			3	0	✓
289	MEMBRANES				✓				✓				✓	3	0	
290	MENTIONED	✓				✓				✓				0	3	
291	METERS				✓			✓		✓				2	1	✓
292	METHANE				✓				✓		✓			3	0	✓
293	METHANOL				✓				✓		✓			3	0	✓
294	METHODOLOGY	✓				✓				✓				0	3	
295	METHODS	✓						✓		✓				1	2	
296	MICRO				✓		✓			✓				2	1	✓
297	MICROSTRIP				✓				✓				✓	3	0	✓
298	MIN	✓				✓						✓		1	2	
299	MINIMIZE	✓				✓				✓				0	3	
300	MINIMIZING	✓				✓				✓				0	3	

301	MIXTURE		✓					✓		✓				2	1	✓
302	MIXTURES		✓					✓		✓				2	1	✓
303	MODAL			✓				✓				✓		3	0	✓
304	MODELED			✓		✓					✓			2	1	✓
305	MODELS			✓				✓			✓			3	0	✓
306	MODES			✓				✓				✓		3	0	✓
307	MODIFIED	✓				✓						✓		1	2	
308	MODULES				✓			✓				✓		3	0	✓
309	MONITOR			✓				✓		✓				2	1	✓
310	MONTHLY	✓				✓				✓				0	3	
311	MOREOVER	✓				✓				✓				0	3	
312	MULTIPATH												✓			
313	NAMELY	✓				✓				✓				0	3	
314	NETWORKS				✓			✓				✓		3	0	✓
315	NEURAL				✓				✓				✓	3	0	✓
316	NEUTRONS				✓			✓		✓				2	1	✓
317	NODES				✓			✓					✓	3	0	✓
318	NORMALIZED		✓			✓						✓		2	1	✓
319	NUSSELT				✓			✓					✓	3	0	✓
320	OBJECTIVE	✓				✓				✓				0	3	
321	OBJECTIVES	✓				✓				✓				0	3	
322	OBSERVATIONS	✓				✓				✓				0	3	
323	OBSERVE	✓				✓				✓				0	3	
324	OBSERVED	✓				✓				✓				0	3	
325	OBSTACLES		✓			✓				✓				1	2	
326	OBTAINED		✓			✓				✓				1	2	
327	OCCURRENCE		✓			✓				✓				1	2	
328	OCCURRENCES		✓			✓				✓				1	2	
329	OCCURS		✓			✓				✓				1	2	

330	ONLINE		✓			✓				✓			1	2	
331	OPERATORS				✓			✓		✓			2	1	✓
332	OPTIMIZE		✓				✓			✓			2	1	✓
333	OPTIMIZED		✓				✓			✓			2	1	✓
334	ORGANIZATIONS		✓					✓		✓			2	1	✓
335	ORTHOGONAL				✓				✓			✓	3	0	✓
336	OUTCOMES			✓			✓			✓			2	1	✓
337	OUTER			✓		✓				✓			1	2	
338	OUTPUTS			✓				✓		✓			2	1	✓
339	OVERALL	✓				✓				✓			0	3	
340	OVERVIEW	✓				✓				✓			0	3	
341	OXIDATION				✓				✓			✓	3	0	✓
342	PACKET			✓					✓		✓		3	0	✓
343	PANELS			✓				✓			✓		3	0	✓
344	PARTIAL		✓			✓					✓		2	1	✓
345	PARTICIPANTS	✓				✓				✓			0	3	
346	PARTICLES			✓				✓			✓		3	0	✓
347	PEAK			✓		✓				✓			1	2	
348	PEAKS			✓		✓				✓			1	2	
349	PERCENT		✓			✓						✓	2	1	✓
350	PERFORM		✓			✓				✓			1	2	
351	PERFORMED		✓			✓				✓			1	2	
352	PERFORMS		✓			✓				✓			1	2	
353	PHANTOM				✓				✓			✓	3	0	✓
354	PHASES			✓			✓			✓			2	1	✓
355	PHONES	✓				✓				✓			0	3	
356	PICKERS			✓				✓			✓		3	0	✓
357	PICKING			✓		✓				✓			1	2	
358	PLATES			✓				✓		✓			2	1	✓

359	PLOT			✓				✓		✓				2	1	✓
360	PLOTS			✓				✓		✓				2	1	✓
361	POLYMER				✓				✓				✓	3	0	✓
362	PORES				✓			✓				✓		3	0	✓
363	POROSITY				✓				✓				✓	3	0	✓
364	POROUS				✓				✓		✓			3	0	✓
365	POWER		✓					✓			✓			3	0	✓
366	PREDICT	✓				✓				✓				0	3	
367	PREDICTED	✓				✓				✓				0	3	
368	PRESENTED	✓				✓				✓				0	3	
369	PRESENTS	✓				✓				✓				0	3	
370	PROCEDURE		✓					✓		✓				2	1	✓
371	PROCEDURES		✓					✓		✓				1	2	
372	PROCESSED		✓			✓				✓				1	2	
373	PROCESSES		✓					✓		✓				2	1	✓
374	PROCESSING		✓					✓		✓				2	1	✓
375	PROCESSOR			✓				✓			✓			3	0	✓
376	PRODUCTS		✓					✓		✓				2	1	✓
377	PROGRAMS			✓				✓		✓				2	1	✓
378	PROPAGATION			✓		✓						✓		2	1	✓
379	PROPOSED		✓			✓				✓				1	2	
380	PROTOCOL		✓					✓		✓				2	1	✓
381	PROVIDER		✓					✓		✓				2	1	✓
382	PROVIDERS		✓					✓		✓				2	1	✓
383	PROVIDES		✓			✓				✓				1	2	
384	PURE			✓		✓				✓				1	2	
385	QUALITATIVE		✓			✓				✓				1	2	
386	QUANTITATIVE		✓			✓				✓				1	2	
387	QUESTIONNAIRE	✓				✓				✓				0	3	

388	QUEUES		✓			✓				✓				1	2	
389	RADIOTHERAPY				✓			✓					✓	3	0	✓
390	RANGES			✓				✓		✓				2	1	✓
391	RANGING			✓				✓		✓				2	1	✓
392	RATED		✓			✓				✓				1	2	
393	RAYS				✓			✓					✓	3	0	✓
394	REACHABILITY			✓				✓		✓				2	1	✓
395	REACHES			✓		✓				✓				1	2	
396	RECOMMENDATIONS	✓				✓				✓				0	3	
397	RECTANGULAR				✓			✓					✓	3	0	✓
398	REDUCE		✓			✓				✓				1	2	
399	REDUCES		✓			✓				✓				1	2	
400	REFERS	✓				✓				✓				0	3	
401	REGARDING	✓				✓				✓				0	3	
402	REGION	✓				✓				✓				0	3	
403	REGRESSION			✓				✓				✓		3	0	✓
404	REINFORCED			✓				✓				✓		3	0	✓
405	RELATED	✓				✓				✓				0	3	
406	RELIABLE		✓			✓				✓				1	2	
407	REMOVING	✓				✓				✓				0	3	
408	REPORTING	✓				✓				✓				0	3	
409	REPRESENT	✓				✓				✓				0	3	
410	REPRESENTS	✓				✓				✓				0	3	
411	REQUIRED	✓				✓				✓				0	3	
412	REQUIREMENT	✓				✓				✓				0	3	
413	REQUIRES	✓				✓				✓				0	3	
414	RESEARCHER	✓				✓				✓				0	3	
415	RESEARCHERS	✓				✓				✓				0	3	
416	RESIDUAL			✓				✓				✓		3	0	✓

417	RESPECTIVELY	✓				✓				✓				0	3	
418	RESULTING	✓				✓				✓				0	3	
419	RESULTS	✓				✓				✓				0	3	
420	REYNOLDS				✓				✓				✓	3	0	✓
421	RISKS		✓					✓		✓				2	1	✓
422	ROUGHNESS			✓				✓		✓				2	1	✓
423	SAMPLES	✓						✓		✓				1	2	
424	SATISFACTION	✓				✓				✓				0	3	
425	SCALES			✓				✓		✓				2	1	✓
426	SCATTERED			✓		✓						✓		2	1	✓
427	SCENARIOS	✓						✓			✓			2	1	✓
428	SCORE			✓		✓				✓				1	2	
429	SCORES			✓		✓				✓				1	2	
430	SELECT	✓				✓				✓				0	3	
431	SELECTING	✓				✓				✓				0	3	
432	SELECTION	✓				✓				✓				0	3	
433	SEQUENCES			✓				✓				✓		3	0	✓
434	SHAPED		✓			✓				✓				1	2	
435	SHAPES		✓			✓				✓				1	2	
436	SHIPPED	✓				✓				✓				0	3	
437	SHOWN	✓				✓				✓				0	3	
438	SHOWS	✓				✓				✓				0	3	
439	SIGNALS			✓				✓		✓		✓		3	0	✓
440	SIGNIFICANT	✓				✓				✓				0	3	
441	SIGNIFICANTLY	✓				✓				✓				0	3	
442	SILICON				✓				✓		✓			3	0	✓
443	SIMULATED		✓					✓			✓			3	0	✓
444	SIMULATION		✓					✓			✓			3	0	✓
445	SIMULATIONS		✓					✓			✓			3	0	✓

446	SIMULTANEOUSLY		✓			✓					✓			2	1	✓
447	SITUATIONS	✓				✓				✓				0	3	
448	SIZES	✓				✓				✓				0	3	
449	SLOTS			✓				✓				✓		3	0	✓
450	SOFTWARE				✓			✓			✓			3	0	✓
451	SOLUTIONS	✓					✓			✓				1	2	
452	SOLVE	✓				✓				✓				0	3	
453	SOLVED	✓				✓				✓				0	3	
454	SOLVING	✓				✓				✓				0	3	
455	SOURCES	✓						✓		✓				1	2	
456	SPACING			✓			✓			✓				2	1	✓
457	SPECIALIZED	✓				✓				✓				0	3	
458	SPECIFIC	✓				✓				✓				0	3	
459	SPECIFICATION	✓						✓		✓				1	2	
460	SPECIFIED	✓				✓				✓				0	3	
461	SPECIMEN	✓					✓						✓	2	1	✓
462	SPECIMENS	✓					✓						✓	1	2	
463	SPECKLE	✓							✓				✓	2	1	✓
464	SPECTRA		✓			✓							✓	2	1	✓
465	SPEEDS		✓					✓					✓	3	0	✓
466	STAGES		✓			✓				✓				1	2	
467	STATIONS		✓					✓		✓				2	1	✓
468	STATISTICS		✓			✓				✓				1	2	
469	STEPS	✓				✓				✓				0	3	
470	STIR			✓				✓			✓			3	0	✓
471	STORED	✓						✓		✓				3	0	✓
472	STRENGTHS	✓						✓		✓				1	2	
473	STRESSES			✓		✓				✓				1	2	
474	STUDENTS	✓				✓				✓				0	3	

475	STUDIED	✓				✓				✓				0	3	
476	STUDIES	✓				✓				✓				0	3	
477	STUDY	✓				✓				✓				0	3	
478	SUB	✓				✓				✓				0	3	
479	SUBSCRIBERS	✓				✓				✓				0	3	
480	SUBSTATIONS	✓						✓					✓	2	1	✓
481	SUITABLE	✓				✓				✓				0	3	
482	SUMMARIZED	✓				✓				✓				0	3	
483	SUMMARY	✓				✓				✓				0	3	
484	SUPERVISION			✓		✓				✓				1	2	
485	SUPERVISOR	✓					✓			✓				1	2	
486	SUPPLIER		✓					✓		✓				2	1	✓
487	SUPPLIERS		✓					✓		✓				2	1	✓
488	SUPPORTS		✓					✓		✓				2	1	✓
489	SURFACES			✓				✓		✓				2	1	✓
490	SUSTAINABILITY		✓				✓			✓				2	1	✓
491	SWITCHES		✓					✓		✓				2	1	✓
492	SYMBOLS			✓			✓			✓				2	1	✓
493	TANKS			✓			✓			✓				2	1	✓
494	TASKS		✓			✓				✓				1	2	
495	TECHNIQUES		✓					✓		✓				2	1	✓
496	TECHNOLOGIES		✓					✓		✓				2	1	✓
497	TELECOMMUNICATION		✓					✓		✓				2	1	✓
498	TEMPERATURES			✓				✓		✓				2	1	✓
499	TERMINALS			✓				✓				✓		3	0	✓
500	TESTED		✓			✓				✓				1	2	
501	TESTING		✓				✓			✓				2	1	✓
502	THESIS	✓				✓				✓				0	3	
503	THICKNESS			✓		✓				✓				1	2	

504	TISSUE			✓				✓				✓		3	0	✓
505	TISSUES			✓				✓				✓		3	0	✓
506	TOOLS				✓			✓		✓				2	1	✓
507	TOPOLOGY				✓				✓				✓	3	0	✓
508	TOTAL	✓				✓				✓				0	3	
509	TOWARD	✓				✓				✓				0	3	
510	TRANSFORM		✓					✓		✓				2	1	✓
511	TRANSMIT		✓					✓					✓	3	0	✓
512	TRANSMITTED		✓					✓					✓	3	0	✓
513	TREND		✓			✓				✓				1	2	
514	TUBES			✓				✓		✓				2	1	✓
515	TUMOR	✓				✓							✓	1	2	
516	TURBINES				✓				✓				✓	3	0	✓
517	TYPES		✓			✓				✓				1	2	
518	TYPICALLY	✓				✓				✓				0	3	
519	ULTRASOUND				✓			✓			✓			3	0	✓
520	UNIVERSITIES	✓				✓				✓				0	3	
521	UNIVERSITY	✓				✓				✓				0	3	
522	UNSAFE	✓					✓			✓				1	2	
523	UPDATE	✓				✓				✓				0	3	
524	USAGE	✓				✓				✓				0	3	
525	USED	✓				✓				✓				0	3	
526	USER	✓				✓				✓				0	3	
527	USERS	✓				✓				✓				0	3	
528	USES	✓				✓				✓				0	3	
529	USING	✓				✓				✓				0	3	
530	UTILIZED		✓			✓				✓				1	2	
531	UTILIZING		✓			✓				✓				1	2	
532	VALIDATION		✓					✓					✓	3	0	✓

533	VALUES			✓		✓				✓				1	2	
534	VARIABLES		✓					✓		✓				2	1	✓
535	VARIANCE		✓				✓					✓		3	0	✓
536	VARIATION		✓				✓					✓		3	9	✓
537	VARIATIONS		✓				✓					✓		3	9	✓
538	VARIES		✓			✓				✓				1	2	
539	VARYING		✓			✓				✓				1	2	
540	VECTORS			✓					✓				✓	3	0	✓
541	VERSUS	✓				✓				✓				1	3	
542	VOLTAGES				✓				✓				✓	3	0	✓
543	WAREHOUSES				✓	✓				✓				1	2	
544	WAVELET				✓			✓					✓	3	0	✓
545	WAVES			✓				✓				✓		3	0	✓
546	WEBSITE	✓				✓				✓				0	3	
547	WEIGHTED	✓				✓				✓				0	3	
548	WEIGHTS	✓						✓		✓				1	2	
549	WIDTH		✓					✓		✓				2	1	✓
550	WIMAX				✓				✓				✓	3	0	✓
551	WIRES			✓				✓			✓			3	0	✓
552	WORKFLOW				✓				✓	✓				2	1	✓
553	WORKPLACE				✓	✓				✓				1	2	
554	YEARLY	✓				✓				✓				0	3	

Appendix B: Teachers' Survey

Instructions

This survey aims to identify the most useful engineering technical vocabulary for engineering students at the university level in Saudi universities. The survey consists of three sections, each containing a set of vocabulary words to be evaluated from most useful to least useful for teaching.

For example, if you believe that knowing the word:

Station is not useful

Apply is Moderately useful

Voltage is Extremely useful

your assessment would be as follows: Station: 1, Apply: 3, Voltage: 5

High-Frequency Technical Vocabulary

NO.	Technical Vocabulary	1	2	3	4	5
1	Drill					
2	Converter					
3	Curves					
4	Spacer					
5	Filters					
6	Carbon					
7	Coolant					
8	Rays					
9	Fibers					
10	Rating					
11	Controller					
12	Feeder					
13	Density					
14	Estimation					
15	Hazards					
16	Barriers					
17	Coupling					
18	Adaptive					
19	Cooling					
20	Emission					
21	Heating					
22	Conductivity					
23	Loading					
24	Loads					
25	Waves					
26	Picker					
27	Machine					
28	Microhardness					
29	Substation					
30	Transmitter					
31	Agents					
32	Application					
33	Steam					
34	Stiffness					
35	Technology					

36	Absorption					
37	Cycle					
38	Boundary					
39	Capacity					
40	Components					
41	Safety					
42	Variation					
43	Approximation					
44	Tracking					
45	Storage					
46	Scaling					
47	Molecular					
48	Carrier					
49	Surfaces					
50	Switching					
51	dimensional					
52	Tool					
53	Patch					
54	Experiment					
55	Structure					
56	Beams					
57	Permittivity					
58	Shield					
59	Quantity					
60	Mechanism					
61	Smoothing					
62	Wires					
63	Architecture					
64	Reactor					
65	Magnetic					
66	Transformer					
67	Absorption					
68	Tubes					
69	Installation					
70	Lean					

Mid-Frequency Technical Vocabulary

NO.	Technical Vocabulary	1	2	3	4	5
1	Ceramic					
2	Turbine					
3	Torque					
4	Electromagnetic					
5	Armchair					
6	Vacuum					
7	Condenser					
8	Pores					
9	Aluminum					
10	Analyzer					
11	Modulation					
12	Iteration					
13	Saline					
14	Optimum					
15	Antenna					
16	Rotor					

17	Pixel					
18	Deformation					
19	Configurations					
20	Amplifier					
21	Warehouse					
22	Velocities					
23	Deposition					
24	Induced					
25	Magnitude					
26	Electron					
27	Entities					
28	Rotational					
29	Threshold					
30	Residual					
31	Silicon					
32	Fleet					
33	Fluid					
34	Particle					
35	Sodium					
36	Solar					
37	Aviation					
38	Axis					
39	Composites					
40	Cylinder					
41	Degradation					
42	Cellular					
43	Feasibility					
44	Discrete					
45	Distillation					
46	Friction					
47	Incubators					
48	Kinetic					
49	Pendant					
50	Amplitude					
51	Attenuation					
52	Compression					
53	Fuzzy					
54	Altitude					
55	Shear					
56	Diameter					
57	Hardware					
58	Voltage					
59	Humidity					
60	Generators					
61	Hollow					
62	Triangular					
63	Robot					
64	Reservoir					
65	Welding					
66	Valves					
67	Corrosion					
68	Hierarchy					
69	Membrane					
70	Fossil					

Low-Frequency Technical Vocabulary

NO.	Technical Vocabulary	1	2	3	4	5
1	Microfiltration					
2	Modularized					
3	Nonadaptation					
4	Convection					
5	Exponential					
6	Speckle					
7	Axial					
8	Ergonomic					
9	Hydroelectric					
10	Impedance					
11	Modal					
12	Reynolds					
13	Zircaloy					
14	Modulus					
15	Topology					
16	Zigzag					
17	Desalination					
18	Outage					
19	Hydropower					
20	Neutron					
21	Penstock					
22	Analog					
23	Doppler					
24	Fourier					
25	Pelton					
26	Epoxies					
27	Graphene					
28	Neutrons					
29	Hydrophobicity					
30	Inverter					
31	Boron					
32	Isothermal					
33	Alloying					
34	Anisotropic					
35	Antiscalant					
36	Wavelet					
37	Macrostructure					
38	Delamination					
39	Detwinned					
40	Linearization					
41	Lognormal					
42	Seawater					
43	Setup					
44	Workflow					
45	Wimax					
46	Microstrip					
47	Biofouling					
48	Cavitating					
49	Radiotherapy					
50	Polarizations					
51	Ultrafiltration					
52	Throughput					
53	Orthogonal					
54	Austenite					
55	Entropy					

56	Circumferentially					
57	Poiseuille					
58	Checksum					
59	Instrumented					
60	Gaussian					
61	Anodization					
62	Tensile					
63	Irradiance					
64	Methanol					
65	Dielectric					
66	Annulus					
67	Dialysis					
68	Demodulator					
69	Bandwidth					
70	Foulant					

Appendix C: Sheet of Information for Participants

Project title: **Investigating knowledge and Usage of Technical Vocabulary in Saudi Engineering Masters' dissertations: A Corpus-Based Study**

Thank you for your interest in this project. Please read this information before deciding whether or not to take part. If you decide to participate, thank you. If you decide not to take part, thank you for considering my request.

Who am I?

My name Budur Alanazi and I am a Doctoral student in Applied Linguistics at Nottingham University. This research project is work towards my thesis.

What is the aim of the project?

This project aims to determine whether students are familiar with understanding technical vocabulary in engineering and lecturers' perceptions of the given technical words. This research has been approved by Nottingham University Human Ethics Committee.

How can you help?

If you agree to participate, you will be asked to complete an online survey. This survey will include questions about technical vocabulary. You have the option to discontinue participation at any time, for any reason. If you choose to withdraw from the study, please contact me before April 2024. In the event of withdrawal, any information provided by you will be securely deleted or returned to you.

What will happen to the information you give?

This research is confidential. This means that the researchers named below will be aware of your identity, but the research data will be aggregated, and your identity will not be disclosed in any reports, presentations, or public documentation. Only my supervisors and I will have access to the survey responses and any associated data. The survey results will be kept securely and destroyed in four years.

Appendix D: Consent Form

Consent to Participate

Project Title: Investigation Knowledge and Use of Technical Vocabulary in Saudi Engineering Masters' Dissertations: A Corpus-based Study

This consent form will be held for 4 years.

Researcher: Budur Alanazi, School of English, Nottingham University

- I have read the Information Sheet, and the project has been explained to me. My questions have been answered to my satisfaction. I understand that I can ask further questions at any time.
- I agree to take part in online tests and survey.

I understand that:

- I may withdraw from this study at any point before 01 April, 2024, without giving any reason, and any information that I have provided will be returned to me or destroyed.
- The information I have provided will be destroyed four (4) years after the research is finished.
- Any information I provide will be kept confidential to the researcher and the supervisor. I understand that the results will be used for a PhD report and a summary of the results may be used in academic reports and/or presented at conferences.
- My name will not be used in reports, nor will any information that would identify me.
- I would like a copy of the transcript of my interview: Yes , NO
- I would like a summary of the findings: Yes , NO

Signature of participant ……………………………….
Name of participant: …………………………………..
Date: ……………………………….
Email address: ………………………….

Appendix E: Condensed Engineering Technical Multiword Units

No.	slots	**Condensed E MWUs**	slots	Freq.
1	in	accredited hospitals		69
2		Accuracy Measures		28
3	of	activated corrosion	product	93
4		activated CRUD		31
5		active antenna		80
6		active power		54
7		after irradiation		29
8		ambient temperature		36
9		analytic Hierarchy		29
10		the steam generator/s	tube	(185+27) 212
11		antenna element/s		(10+21) 31
12		antenna array/s		(52+12) 64
13		antenna parameters	performance	49
14		array antenna		(32+5) 37
15		array patch		24
16		Artificial Neural	network	54
17		aspect ratio		45
18	Width Minimum	Average Maximum		45
19	shows the/ the	average number	waiting/ of customers/ of customers out	139
20		average recognition		34
21		average value		28
22		bar code		31
23		barometric desalination		36
24		base case		50
25	the	base metal		35
26	the	base station/s		(93+53) 146
27		battery storage		26
28	the	Block diagram		123
29		boron compound		26
30		boundary condition/s		(45+94) 139
31		boundary layer		31
32		brake disc		23
33		brake power		25
34		brake torque		22
35	the/ laminar	burning velocity		142
36		bus voltage/s		(50+29) 79
37		business incubators		(12+48) 60
38		business processes		24
39		Carbon Ceramic		42
40		carbon dioxide		30
41		carbon fiber		39
42		Catalytic activity	of	54
43		channel estimation		37
44	the/in	circuit breaker/s		(223+42)265
45		circuit diagram		29
46		circular polarization		44
47		circular polarized		56
48		classification accuracy		36
49		code length		21
50		coefficient of variation		28
51		communication system/s		(37+42) 79
52		Comparison Matrix		35
53		composite material/s		(42+88) 130

54	the	compressive strength		109
55		concentration polarization		29
56		conceptual design		24
57		control system		34
58		convection heat		31
59		coolant activity	due to	37
60		coolant flow		25
61		coolant specific	activity	46
62		cooling water		92
63		core strength		22
64		correction factor/s		(20+23) 43
65		correlation coefficient		32
66	activated	corrosion product/s	activity	(90+58) 148
67		corrosion rate		57
68		count rate		44
69		crack depths		25
70		crack position		23
71		crack size/s		(24+20) 44
72		critical success	factor	40
73		cross validation		63
74		cubic meters		23
75		cutting speed		28
76		cycle counting		81
77		cycle time		27
78		data transmission		29
79		decision matrix		48
80		delamination factor		25
81		delay cost		40
82		delay spread		27
83		delay time		56
84		dependent variable		29
85		desalination system/s		(40+12) 52
86		design parameters		24
87		design variables		40
88		DG system		26
89		dielectric constant		47
90		dielectric substrate		52
91		diesel fuel		41
92		diesel generator		95
93		different crack		40
94		different images	with	44
95		different locations		20
96		different velocities		20
97		differential equations		24
98		diffusion filter		40
99		digital transformation		71
100		discrete event		42
101		distillation process		20
102		distribution network		39
103	the	distribution system		50
104		dose distribution		25
105		dose enhancement	factor/ factor due to	76
106	the	draft tube		50
107		drill bit		42
108	the	drilling process		35
109		dual fuel		38
110		Dynamic Programming		97
111		e-commerce system		28

112	mean	effective pressure		25
113	the/Relative	efficiency results		133
114		electric power		66
115		electrical energy	consumption	38
116		electrical load/s		(26+21) 47
117		Electricity Company		48
118		electricity consumption		35
119		electricity demand		97
120		electricity generation		33
121		Element Table		29
122		energy consumption		76
123		energy demand		24
124		energy density		56
125		energy resource/s		(45+11) 56
126		energy source/s		(25+99)124
127		energy system/s		(57+14) 71
128		engine performance		25
129		engine speed		24
130		enhancement factor	due/due to	71
131		epithermal neutron	beam	49
132		equilibrium corrosion		27
133		equipment effectiveness		26
134		error rate		37
135		experimental data		32
136		experimental results		66
137		experimental study		28
138		experimental work		36
139		fading channel		35
140		failure rate		28
141		fault tree		97
142		feature extraction		72
143		feature vector		24
144		feed channel		34
145		feed line		48
146		feed side		31
147		feed solution		31
148		feed stream		25
149		feed temperature		61
150		feed water		62
151	hollow	fibre membranes		32
152		filtered image		51
153		Filtering results		32
154		filtering techniques		35
155		finite element		66
156		fishing weights		28
157		flight data		34
158		flight delay		28
159		flow analysis		35
160		flow chart		39
161		flow velocity		38
162		fluid flow		29
163		flux decline		23
164	pure	forced convection		35
165		fossil fuels		36
166		frequency band/s		(46+31) 77
167		frequency domain		50
168		frequency response		66
169		friction factor		97

170		friction stir		53
171		the front panel		27
172		FSW process		21
173		fuel cell		27
174		fuel consumption		38
175		gas circuit	breaker	41
176		generator tubes		20
177		genetic algorithm		28
178		Glass Fiber		22
179		governing equations		28
180		grid load	following	34
181		grid power		32
182		Grid connected	PV	95
183		(the) ground plane		100
184		Gunn diode		34
185	Average	Half Width	Minimum	28
186		heat exchanger		103
187	constant/wall	heat flux	boundary conditions/at/case	160
188		heat loss		33
189		hidden layer/s		(943+47) 90
190		Hierarchy Process		33
191		high frequency		41
192		high speed		65
193		high temperature		57
194		high voltage		74
195		hollow fibre		58
196		Hybrid Median		40
197		hybrid renewable		33
198		hybrid system		70
199		hydroelectric power		51
200		hydropower plant		58
201	The positive	ideal solution		91
202		image quality		45
203		images obtained		25
204		impedance bandwidth		49
205		impulse response		29
206		incident angles		27
207		injection timing		35
208		inlet temperature		61
209		inlet velocity/velocities		(30+54) 84
210		input data		45
212	the	input impedance		34
213	(**two** tables for /tables for)	input orientation	and output/and	27
214		Input Oriented	target/ target value	28
215		input signal/s		(20+45) 65
216		installed capacity		28
217		interface unit		22
218		internal variables		31
219		Interpretive Structural	Modelling	26
220		inventory accuracy		41
221		iodine concentration		23
222		keV photon	for	36
223		kinetic energy		62
224		laminar flow		28
225		large scale		83
226	of	lean manufacturing		79
227		Lee filter		28

228	the	license plate/s	recognition	116+71) 187
229	the	load demand		96
230		load distribution		68
231		load flow		88
232		load following		42
233		load forecasting		88
234		load profile		25
235		loading level		72
236		Loc bus		24
237		low noise		24
238		low temperature		30
239		low voltage		55
240		MAC layer		29
241	the	magnetic field		107
242		main components		35
243		management system/s		73
245		manufacturing process		26
246		mass flux		44
247		mass transfer		70
248		mathematical model		63
249		Maximum Average		52
250		maximum temperature		29
251	the	MD process		41
252		mean temperature		49
253		mechanical energy		23
254		mechanical power		27
255		mechanical properties		52
256		median filter		63
257		medical imaging		39
258		Medium Enterprises		44
259		medium voltage		28
260	in/contact/ direct contact	membrane distillation		195
261		membrane pores		20
262		the membrane	channel/surface	426
263	shape	memory alloys		49
264		merging filter		27
265	in	meters equation		22
266	of	microstrip antenna/s		(98+54) 152
267		middle defect		28
267		MIMO system		68
268		mixed convection		28
269		mobile phone/s		(57+43) 100
270		mobile WiMAX		76
271		modal participation		61
272		modulation scheme/s		(11+24) 35
273		Moving Average		43
274		multipath fading		29
275		mutual coupling		51
276	the	natural convection		52
277	the	negative ideal	solution	37
278		network output		22
279	artificial	neural network		150
280		neutron beam		70
281		neutron flux		74
282		noise reduction		36
283	the	Nusselt number/s		(166+19) 185
284		OFDM system		38
285		operating conditions		54

286		operating parameters		34
287		operation cost		71+6) 77
288	the	optimal solution		82
289	the total	Optimum Value		48
290		organizational structure		24
291	Output	Oriented Target	values	54
292		original image		50
293		outer radius		29
294	orientation and	output orientation	respectively	27
295		overall average		20
296		parameters performance		33
297	the	parameters of		131
298		partial oxidation		31
299		particle size		22
300		passive antenna		34
301		path loss		35
302	number of	Payment Machine		113
303		PDZ domain/s		86+80) 166
304		peak load		38
305		penetration level/s		(15+21)
306		per unit		32
307		performance indicators		(8+38)46
308		permeate flow		28
309	the	permeate flux		140
310	the	permeate side		26
311		phase transformation		24
312		photovoltaic system		29
313		physical layer		42
314		picking process		44
315		pin fins		40
316		plate recognition		31
317		polarized radiation	pattern	22
318	for	POM reaction		44
319		pore size		51
320		porous layer		25
321		positive impact		23
322		power consumption		38
322		power factor		54
324		power flow		45
325		power generation		115
326		power grid		27
327		power network		40
328		power sources		23
329		power station/s		45
330	in the/the	power system/s	network	(234+41) 275
331	the	pressure drop		111
332		preventive maintenance		36
333	the	primary coolant	activity/ specific/ specific activity	133
334		processing elements		40
335		processing parameters		25
336		produce electricity		22
337		product activity		24
338		production capacity		31
339		programming language		23
340	the	proposed system		49
341		pure diesel		26
342		push-out delamination		27

343		PV array		(44+5) 49
344		PV cell/s		(22+21) 43
345		PV module/s		22
346		PV penetration		35
347		PV plant		36
348	the	PV system/s		(197+67) 244
349		Quality test	for formula	39
350		radial direction		29
351		radiating patch		27
352	the	radiation pattern/s		(171+56)227
353		raw material/s		(22+31) 53
354		Rayleigh number		53
355		RBF kernel		37
356		reachability matrix		50
357		reaction stages		21
358		reaction turbine		22
359		Reactive Power		134
360		real system		25
361	the	received signal		59
362		recognition system		24
363		rectangular patch	antenna	40
364		reflection coefficient		23
365		remote areas		32
366		Renewable Energy	sources	254
367		reporting system		25
368		research reactor		46
369		residual stress		40
370	the	resonant frequency		47
371		respiratory motion		47
372	the	Reynolds number		165
373		risk factor/s		(9+24) 33
374		room temperature		29
375		root cause/s		(66+39) 105
376		rotation speed		52
377		rotational speed		52
378		rotor angle		27
379		safety culture		109
380		safety management		37
381		sales executives		52
382		Satisfaction percentage		69
383		saturation value		36
384		Scatter plot	of	28
385	SF6	circuit breaker		223
386		spherical shaped	module/endoscope	54
387		short circuit	current	43
388		SiC particles		26
389		signal strength		23
390		significant impact		24
391		slotted angle		31
392		SMA wires		31
393		small enterprises		20
394		solar cell/s		(49+66) 115
395		solar irradiance		36
396		solar panel		21
397		solar photovoltaic		50
398		solar radiation		84
399	the	spacer filament/s		15+45) 60
400	primary coolant	specific activity		125

401		specific speed		21
402		speckle noise		79
403		Spectral Efficiency	of/for	86
404		spindle speed		42
405	Mean	Square Error		39
406		squared error		25
407		stability analysis		29
408		standard cylinder		21
409		standard deviation		87
410		statistical analysis		29
411	the	steady state		138
412		stir welding		33
413		stop band		24
414		storage system		40
415		Structural Modelling		24
416		substation building	cost	47
417	the/green	supply chain	management	164
418		Support Vector	Machine	43
419		surface area		63
420		surface roughness		40
421		system components		37
422		system configuration		30
423		system consists of		25
424		system design		40
425		system network		22
426	the	system performance		59
427		system stability		23
428	a	system used	in	26
429	Input Oriented/ Actual Value/Value	Target Value/s	Potential/improvement	55
430		teach pendant		76
431		temperature contour		25
432		temperature difference		52
433		temperature distribution		23
433		temperature gradient/s		(29+9) 38
434		temperature mapping		31
435	the	temperature polarization		64
436		temperature profile		35
437		temperature range		23
438		temperature rise		25
439	the/ ultimate	tensile strength		118
440		thermal camera		25
441		thermal column		40
442		thermal conductivity		84
443		thermal efficiency		34
444		thermal flux		30
445		thermal neutron		29
446		thermal power		20
447	the	thrust force	values	180
448		time domain		45
449		time step		30
450		tool rotation		29
451		total consumption		27
452	the	Total Optimum	value	46
453		tracking system/s		(28+28) 56
454		transfer coefficient		88
455		transfer function		45

456		Transfer Learning		34
457		transfer rate		25
458		transient stability		38
459		traverse speed		42
460		Triangular Patch	Antennas Arrays	53
461		Triangular web		34
462		turbine power		26
463		turbine speed		26
464		ultimate tensile		36
465		ultrasound image/s		(26+36) 62
466		ultrasound imaging		31
467		unit cell		29
468		user interface		30
469		utility grid		29
470		UWB antenna		55
471		Vapour/vapor pressure		(28+37) 65
472		variable properties		34
473		Vector Machine		30
474		velocity profile		41
475		vertical annulus		28
476	mode of	Vibration Weak		20
477		voltage east		36
478		voltage stability		62
479		voltage value		28
480		wall temperature		64
481		water inlet	temperature	31
482		water temperature		28
483		water vapor/vapour		(25+9) 34
484		wavelet function/s		(29+27) 56
485		wavelet transform		48
486		weather conditions		36
487		web element		21
488		welding process		31
489		welding speed		30
490		wind speed		53
491		wind turbine		60
492		wireless communication/s		(47+21) 68
493		work conditions		69
494		work environment		47
495		working conditions		25
496	show the	average scores		29
497	the/show the	average waiting time		53
498	the	core scale		63
499	the	electrical power		186
500	the/feed/ mass/water	flow rate/s		(269+40) 309
501	the	heat transfer	*coefficient/rate/in*	370
502		the natural frequencies		370
503	the	output layer		52
504	the/polarized a microstrip/microstrip/shaped	Patch Antenna/s		(411+92)503
505	The	Qmatic machine		27
506		the solution depicted		27
507		the transmission line	effect/of	109
508		the signal	of	150
509	the	transmitted signal		42
510	the	average of the total		20
511		brake thermal efficiency		21
512		Centre Frequency BW	Remark	22

513		During the drilling	of	31
514		elements in dimension		24
515		feed and permeate		26
516		Frequency BW Remark		22
517	occupational	health and safety		91
518		in the condenser		22
519		increase in operation		25
520		Input/s and output/s		(25+24) 49
521		internal and external		26
522		Interpretive Structural Modelling		24
523		Membrane feed channel		21
524	Model Factor Actual	Value Target	actual	56
525	Input Oriented/Output Oriented	Target Values for/Potential	Improvement Model	82
526		Mode of Vibration		28
527		number of cycles		50
528		number of substations		68
529		on the surface		29
530	out from/on	the system		979
531		probability of error		28
532		products and services		27
533		safety and health		39
534		shows the average	scores	98
535		strength of concrete		27
536		test for formula		29
537	the	electric field		39
538		the reactor	core	169
539	The	SMART reactor		76
540	the/of/this	proposed antenna		101
541		transmitter and receiver		27
542		increasing of sales executives		20
543		irradiation and evacuation	from	24

Appendix F: Engineering Technical Vocabulary

1. ABSORPTION	2. ACCREDITED	3. ACCURACY
4. ACTIVATED	5. ACTIVE	6. ACTIVITY
7. ACTUATIO	8. ACTUATOR	9. ADAPTIVE
10. ADDING	11. ADOPTION	12. AFFECT
13. AFFECTING	14. AFFECTS	15. AGENTS
16. AGGREGATE	17. AIRCRAFT	18. AIRCRAFTS
19. AIRLINE	20. AIRLINES	21. ALGORITHM
22. ALGORITHMS	23. ALLOWS	24. ALLOY
25. ALLOYING	26. ALLOYS	27. ALTERNATIVES
28. ALTITUDE	29. ALUMINUM	30. AMBIENT
31. AMPLIFIER	32. AMPLITUDE	33. ANALOG
34. ANALYSIS	35. ANALYTICAL	36. ANALYZE
37. ANALYZED	38. ANALYZER	39. ANALYZING
40. ANGLE	41. ANGLES	42. ANISOTROPIC
43. ANNULUS	44. ANODIZATION	45. ANTENNA
46. ANTENNAS	47. ANTISCALANT	48. ANTISCALANTS
49. APPLICABLE	50. APPLICATION	51. APPLICATIONS
52. APPLIED	53. APPLYING	54. APPROACHES
55. APPROXIMATELY	56. APPROXIMATION	57. ARCHITECTURE
58. ARMCHAIR	59. ARRAY	60. ARRAYS
61. ARTIFICIAL	62. ASYMPTOTES	63. ATOMIC
64. ATTENUATION	65. ATTRIBUTES	66. AUSTENITE
67. AUTOMATION	68. AVAILABILITY	69. AVERAGE
70. AVERAGING	71. AVIATION	72. AXIAL
73. AXIS	74. BAND	75. BANDS
76. BANDWIDTH	77. BAROMETRIC	78. BARRIER
79. BARRIERS	80. BASE	81. BATTERIES
82. BATTERY	83. BEAM	84. BEAMS
85. BENDING	86. BIAS	87. BIOFOULING
88. BLOCK	89. BORON	90. BOUNDARY
91. BRAKE	92. BREAKER	93. BREAST
94. BRINE	95. BURNING	96. BUS
97. BUSES	98. BUTTON	99. CABLE
100. CALCULATION	101. CALCULATIONS	102. CAMERA
103. CAPABILITIES	104. CAPABILITY	105. CAPACITY
106. CAPSULE	107. CAPTURE	108. CARBON
109. CARRIER	110. CARRIERS	111. CASTING
112. CATALYST	113. CATALYSTS	114. CATALYTIC
115. CATEGORIES	116. CATEGORY	117. CAVITATING
118. CELL	119. CELLULAR	120. CENTER
121. CERAMIC	122. CHAIN	123. CHANNEL
124. CHANNELS	125. CHART	126. CHECKSUM
127. CHEMICAL	128. CIPHERTEXT	129. CIRCUIT
130. CIRCULAR	131. CIRCUMFERENTIALLY	132. CLASSIFICATION
133. CLASSIFIER	134. CODE	135. CODES
136. CODING	137. COEFFICIENT	138. COEFFICIENTS
139. COLLOIDS	140. COLOR	141. COLUMN
142. COMBINATION	143. COMBINATIONS	144. COMBINING
145. COMBUSTION	146. COMMUNICATIONS	147. COMPACT
148. COMPARISON	149. COMPONENT	150. COMPONENTS
151. COMPOSITE	152. COMPOSITES	153. COMPOSITION
154. COMPOUND	155. COMPRESSION	156. COMPRESSIVE

157. COMPUTATION	158. COMPUTATIONAL	159. COMPUTED
160. CONCENTRATION	161. CONCLUSIONS	162. CONCRETE
163. CONDENSER	164. CONDITIONS	165. CONDUCTIVITY
166. CONFIGURATION	167. CONFIGURATIONS	168. CONNECT
169. CONNECTED	170. CONSTANT	171. CONSTRAINTS
172. CONSTRUCT	173. CONSUMPTION	174. CONTINUOUS
175. CONTROLLER	176. CONVECTION	177. CONVECTIVE
178. CONVENTIONAL	179. CONVERSION	180. CONVERT
181. CONVERTER	182. COOLANT	183. COOLING
184. COORDINATION	185. COPPER	186. CORE
187. CORRECTION	188. CORRELATION	189. CORRELATIONS
190. CORROSION	191. COSINE	192. COST
193. COUNT	194. COUNTING	195. COUPLED
196. COUPLING	197. COVERAGE	198. CRACK
199. CRACKS	200. CRITERIA	201. CRITERION
202. CRITICAL	203. CRYPTION	204. CURRENT
205. CURVE	206. CURVES	207. CYCLE
208. CYCLES	209. CYLINDER	210. DATA
211. DEAERATOR	212. DECOMPOSITION	213. DEFECT
214. DEFECTS	215. DEFORMATION	216. DEGRADATION
217. DELAMINATION	218. DELAY	219. DELAYS
220. DEMAND	221. DEMODULATOR	222. DENSITY
223. DEPOSITION	224. DEPTH	225. DESALINATION
226. DESIGN	227. DESIGNING	228. DESIRED
229. DETECTION	230. DETECTOR	231. DETWINNED
232. DEVIATION	233. DEVICE	234. DIAGRAM
235. DIALYSIS	236. DIAMETER	237. DIAMETERS
238. DIELECTRIC	239. DIESEL	240. DIFFUSION
241. DIGITAL	242. DIMENSION	243. DIMENSIONAL
244. DIMENSIONS	245. DIODE	246. DIRECTION
247. DIRECTIVITY	248. DISC	249. DISCRETE
250. DISCRETIZED	251. DISPERIVE	252. DISPLACEMENT
253. DISTANCE	254. DISTILLATE	255. DISTILLATION
256. DISTORTION	257. DISTRIBUTED	258. DISTRIBUTION
259. DIVERSITY	260. DIVIDED	261. DOMAIN
262. DOMAINS	263. DOPPLER	264. DOSE
265. DOWNSTREAM	266. DRILL	267. DRILLING
268. DRIVEN	269. DUAL	270. DURATION
271. DYNAMIC	272. DYNAMICS	273. EFFECT
274. EFFECTIVE	275. EFFECTIVENESS	276. EFFICIENCY
277. ELECTRIC	278. ELECTRICAL	279. ELECTRICITY
280. ELECTROMAGNETIC	281. ELECTRON	282. ELECTRONIC
283. ELEMENT	284. ELEMENTS	285. ELEVATION
286. EMISSION	287. EMISSIONS	288. ENERGIES
289. ENERGY	290. ENGINEERING	291. ENHANCEMENT
292. ENTERPRISES	293. ENTITIES	294. ENTITY
295. ENTROPY	296. ENVIRONMENT	297. EPITHERMAL
298. EPOXIES	299. EQUALIZATION	300. EQUALIZER
301. EQUATION	302. EQUATIONS	303. EQUIPMENT
304. ERGONOMIC	305. EROSION	306. ERROR
307. ERRORS	308. ESTIMATE	309. ESTIMATION
310. EVALUATION	311. EVAPORATION	312. EVAPORATOR
313. EXCEED	314. EXCHANGER	315. EXECUTIVES
316. EXIT	317. EXPERIMENT	318. EXPERIMENTAL
319. EXPERIMENTALLY	320. EXPERIMENTS	321. EXPONENTIAL

322. EXPOSURE	323. EXTERNAL	324. EXTRACTED
325. EXTRACTION	326. FABRICATED	327. FACTOR
328. FADING	329. FAILURE	330. FAILURES
331. FATIGUE	332. FAULT	333. FAULTS
334. FEASIBILITY	335. FEASIBLE	336. FEATURE
337. FEATURES	338. FEED	339. FEEDER
340. FIBER	341. FIBERS	342. FIGURE
343. FILTER	344. FILTERED	345. FILTERING
346. FILTERS	347. FINITE	348. FINS
349. FIXED	350. FLAME	351. FLEET
352. FLEXIBILITY	353. FLEXIBLE	354. FLIGHT
355. FLOOD	356. FLOW	357. FLUID
358. FLUX	359. FOCUSING	360. FOLLOWING
361. FORECAST	362. FORECASTING	363. FORECASTS
364. FORMULA	365. FORMULATION	366. FOSSIL
367. FOULANT	368. FOULING	369. FOURIER
370. FRACTION	371. FREQUENCIES	372. FREQUENCY
373. FRICTION	374. FUEL	375. FUELS
376. FUNCTION	377. FUNCTIONS	378. FUZZY
379. GAIN	380. GASES	381. GAUSSIAN
382. GENERATE	383. GENERATED	384. GENERATING
385. GENERATION	386. GENERATOR	387. GENERATORS
388. GEOMETRY	389. GRADIENT	390. GRAPH
391. GRAPHENE	392. GRAPHICAL	393. GRID
394. HANDLING	395. HARDNESS	396. HARDWARE
397. HARMONIC	398. HARMONICS	399. HAZARDS
400. HEAT	401. HEATING	402. HEIGHT
403. HIERARCHY	404. HOLLOW	405. HORIZONTAL
406. HUMIDITY	407. HYBRID	408. HYDRAULIC
409. HYDRO	410. HYDROELECTRIC	411. HYDROGEN
412. HYDROPHOBICITY	413. HYDROPOWER	414. IDENTIFICATION
415. IMAGE	416. IMAGES	417. IMAGING
418. IMPACT	419. IMPEDANCE	420. IMPLEMENTATION
421. INCUBATORS	422. INDICATOR	423. INDICATORS
424. INDUCED	425. INFRASTRUCTURE	426. INITIAL
427. INJECTION	428. INLET	429. INNOVATION
430. INPUT	431. INPUTS	432. INSTALLATION
433. INSTALLED	434. INSTRUMENTED	435. INTAKE
436. INTEGRAL	437. INTEGRATED	438. INTEGRATION
439. INTENSITY	440. INTERACTION	441. INTERFACE
442. INTERFERENCE	443. INTERNAL	444. INTERRUPTION
445. INTERVAL	446. INVENTORY	447. INVERSE
448. INVERTER	449. INVERTERS	450. IRRADIANCE
451. IRRADIATION	452. ISOLATION	453. ISOTHERMAL
454. ITERATION	455. ITERATIONS	456. KERNEL
457. KINETIC	458. LABOR	459. LAMINAR
460. LAPLACIAN	461. LAYER	462. LAYERS
463. LAYOUT	464. LEADS	465. LEAN
466. LENGTH	467. LINEAR	468. LINEARIZATION
469. LIQUID	470. LOAD	471. LOADED
472. LOADING	473. LOADS	474. LOCALIZATION
475. LOCATION	476. LOCATIONS	477. LOGIC
478. LOGNORMAL	479. LOOP	480. LOOPS
481. LOW	482. LOWER	483. LUMINANCE
484. MACHINE	485. MACROGRAPHS	486. MACROSTRUCTURE

487. MAGNET	488. MAGNETIC	489. MAGNITUDE
490. MAHALANOBIS	491. MAINTENANCE	492. MANUAL
493. MANUALLY	494. MANUFACTURING	495. MAPPING
496. MARGIN	497. MARTENSITE	498. MARTENSTIC
499. MASS	500. MASSES	501. MATCHING
502. MATERIAL	503. MATERIALS	504. MATHEMATICAL
505. MATRIX	506. MAX	507. MAXIMUM
508. MEASURED	509. MEASUREMENT	510. MEASURING
511. MECHANICAL	512. MECHANISM	513. MEDIAN
514. MEDIUM	515. MEMBRANE	516. MEMBRANES
517. METER	518. METERS	519. METHANE
520. METHANOL	521. METHOD	522. MICRO
523. MICROFILTRATION	524. MICROHARDNESS	525. MICROSTRIP
526. MICROSTRIPS	527. MINIMUM	528. MIXING
529. MIXTURE	530. MIXTURES	531. MOBILE
532. MODAL	533. MODE	534. MODEL
535. MODELED	536. MODELING	537. MODELS
538. MODES	539. MODIFICATION	540. MODULARIZED
541. MODULATION	542. MODULE	543. MODULES
544. MODULUS	545. MOLECULAR	546. MOMENTUM
547. MONITOR	548. MONITORING	549. MOTION
550. MULTIPLE	551. NEGATIVE	552. NETWORK
553. NETWORKING	554. NETWORKS	555. NEURAL
556. NEUTRON	557. NEUTRONS	558. NODES
559. NOISE	560. NOMINAL	561. NONADAPTATION
562. NONLINEAR	563. NORMALIZED	564. NUCLIDES
565. NUMBER	566. NUMERICAL	567. NUSSELT
568. OCCUPATIONAL	569. OPERATING	570. OPERATION
571. OPERATIONAL	572. OPERATIONS	573. OPERATOR
574. OPERATORS	575. OPTIMAL	576. OPTIMIZATION
577. OPTIMIZE	578. OPTIMIZED	579. OPTIMUM
580. ORGANIZATION	581. ORGANIZATIONAL	582. ORGANIZATIONS
583. ORIENTATION	584. ORIENTED	585. ORTHOGONAL
586. OUTAGE	587. OUTCOMES	588. OUTPUT
589. OUTPUTS	590. OXIDATION	591. OXIDE
592. PACKET	593. PANEL	594. PANELS
595. PARALLEL	596. PARAMETER	597. PARAMETERS
598. PARTIAL	599. PARTICLE	600. PARTICLES
601. PARTS	602. PASSIVE	603. PATCH
604. PATH	605. PATTERN	606. PELTON
607. PENDANT	608. PENETRATION	609. PENSTOCK
610. PERCENT	611. PERCENTAGE	612. PERFORMANCE
613. PERIODIC	614. PERMEATE	615. PERMITTIVITY
616. PHANTOM	617. PHASE	618. PHASES
619. PHASORS	620. PHOTON	621. PHOTOVOLTAIC
622. PICKER	623. PICKERS	624. PIPE
625. PIXEL	626. PLANE	627. PLANT
628. PLASTICIZED	629. PLATE	630. PLATES
631. PLOT	632. PLOTS	633. POISEUILLE
634. POLARIZATION	635. POLARIZATIONS	636. POLARIZED
637. POLARIZER	638. POLYMER	639. PORE
640. PORES	641. POROSITY	642. POROUS
643. POWER	644. PREDICTION	645. PRESSURE
646. PRIMARY	647. PROBABILITY	648. PROBE
649. PROCEDURE	650. PROCEDURES	651. PROCESS

652. PROCESSES	653. PROCESSING	654. PROCESSOR
655. PRODUCT	656. PRODUCTIVITY	657. PRODUCTS
658. PROFILE	659. PROGRAM	660. PROGRAMMING
661. PROGRAMS	662. PROPAGATION	663. PROPERTIES
664. PROPORTIONAL	665. PROTOCOL	666. PROVIDER
667. PROVIDERS	668. PULSE	669. PUMP
670. QUALITY	671. QUANTITY	672. QUEUE
673. RADIAL	674. RADIATION	675. RADIOTHERAPY
676. RADIUS	677. RAMPING	678. RANDOM
679. RANGES	680. RANGING	681. RANKING
682. RATE	683. RATING	684. RATIO
685. RAY	686. RAYS	687. REACHABILITY
688. REACTION	689. REACTIVE	690. REACTOR
691. RECEIVER	692. RECOGNITION	693. RECTANGULAR
694. REDUCING	695. REDUCTION	696. REGRESSION
697. REINFORCED	698. RELATIVE	699. RELIABILITY
700. REMOTE	701. RENEWABLE	702. REQUIREMENTS
703. RESEARCH	704. RESERVOIR	705. RESIDUAL
706. RESOLUTION	707. RESONANCE	708. RESONANT
709. RESOURCE	710. RETRANSMISSION	711. REYNOLDS
712. RISKS	713. ROBOT	714. ROOT
715. ROTATING	716. ROTATION	717. ROTATIONAL
718. ROTOR	719. ROUGHNESS	720. ROUTING
721. RUNNER	722. SAFETY	723. SALINE
724. SAMPLE	725. SAMPLING	726. SATURATION
727. SCALE	728. SCALES	729. SCALING
730. SCAN	731. SCATTER	732. SCATTERED
733. SCATTERING	734. SCENARIO	735. SCENARIOS
736. SCHEDULE	737. SCHEDULING	738. SCHEMATIC
739. SEASONAL	740. SEAWATER	741. SEGMENT
742. SEIZE	743. SELECTED	744. SELECTIVITY
745. SENSITIVITY	746. SEPARATION	747. SEQUENCE
748. SEQUENCES	749. SETUP	750. SHAPE
751. SHEAR	752. SHIELD	753. SHIELDING
754. SHIFT	755. SIGNAL	756. SIGNALS
757. SILICON	758. SIMILARITY	759. SIMULATE
760. SIMULATED	761. SIMULATION	762. SIMULATIONS
763. SIMULATOR	764. SIMULTANEOUSLY	765. SIZE
766. SLOT	767. SLOTS	768. SLOTTED
769. SMART	770. SMOOTHING	771. SODIUM
772. SOFTWARE	773. SOLAR	774. SOLUTION
775. SOLUTIONS	776. SOURCE	777. SPACER
778. SPACING	779. SPATIAL	780. SPECIFICATIONS
781. SPECIMEN	782. SPECKLE	783. SPECTRA
784. SPECTRAL	785. SPECTRUM	786. SPEED
787. SPEEDS	788. SPHERICAL	789. SPINDLE
790. SPREADING	791. STABILITY	792. STANDARD
793. STATIC	794. STATION	795. STATIONS
796. STATISTICAL	797. STEADY	798. STEAM
799. STEP	800. STIFFNESS	801. STIR
802. STOICHIOMETRIC	803. STORAGE	804. STORED
805. STRAIN	806. STRATEGIC	807. STREAM
808. STRENGTH	809. STRESS	810. STRUCTURAL
811. STRUCTURE	812. SUBSTATION	813. SUBSTATIONS
814. SUBSTITUTION	815. SUBSTRATE	816. SULFATE

817. SUPERPLASTICITY	818. SUPPLIER	819. SUPPLIERS
820. SUPPLY	821. SUPPORTS	822. SURFACE
823. SURFACES	824. SUSTAINABILITY	825. SUSTAINABLE
826. SWITCH	827. SWITCHES	828. SWITCHING
829. SYMBOL	830. SYMBOLS	831. SYSTEM
832. SYSTEMS	833. TABLE	834. TABLES
835. TAILRACE	836. TANKS	837. TARGET
838. TECHNICAL	839. TECHNIQUE	840. TECHNIQUES
841. TECHNOLOGIES	842. TECHNOLOGY	843. TELECOMMUNICATION
844. TEMPERATURE	845. TEMPERATURES	846. TEMPORAL
847. TENSILE	848. TERMINAL	849. TERMINALS
850. TEST	851. TESTING	852. THEORETICAL
853. THERMAL	854. THERMOMETRY	855. THERMOSET
856. THRESHOLD	857. THROUGHPUT	858. THRUST
859. THYRISTOR	860. TISSUE	861. TISSUES
862. TOOL	863. TOOLS	864. TOPOLOGY
865. TORQUE	866. TRACKING	867. TRACTION
868. TRANSFER	869. TRANSFORM	870. TRANSFORMATION
871. TRANSFORMER	872. TRANSIENT	873. TRANSMISSION
874. TRANSMIT	875. TRANSMITTED	876. TRANSMITTER
877. TRANSPORTATION	878. TRIANGULAR	879. TUBE
880. TUBES	881. TURBINE	882. TURBINES
883. TYPE	884. ULTRAFILTRATION	885. ULTRASOUND
886. UNIFORM	887. UNIT	888. UTILITY
889. UTILIZATION	890. VACUUM	891. VALIDATION
892. VALUE	893. VALVE	894. VALVES
895. VAPOR	896. VARIABLE	897. VARIABLES
898. VARIANCE	899. VARIATION	900. VARIATIONS
901. VECTOR	902. VECTORS	903. VELOCITIES
904. VELOCITY	905. VERTICAL	906. VIBRATION
907. VOLTAGE	908. VOLTAGES	909. VOLUME
910. VORTEX	911. VORTICITY	912. WAREHOUSE
913. WATER	914. WAVE	915. WAVELET
916. WAVELETS	917. WAVES	918. WEB
919. WEIGHT	920. WELDING	921. WIDTH
922. WIMAX	923. WIND	924. WINGLETS
925. WIRELESS	926. WIRES	927. WORKFLOW
928. ZIGZAG	929. ZIRCALOY	930. ZONE

Appendix G: Engineering Technical Vocabulary Identified by the Raters

Words identified by the raters		
High-frequency Vocabulary	**Mid-Frequency Vocabulary**	**Low-Frequency Vocabulary**
Dill	Ceramic	Microfiltration
Converter	Turbine	Modularized
Curves	Torque	Nonadaptation
Spacer	Electromagnetic	Convection
Filters	Armchair	Exponential
Carbon	Vacuum	Speckle
Coolant	Condenser	Axial
Rays	Pores	Ergonomic
Fibers	Aluminum	Hydroelectric
Rating	Analyzer	Impedance
Controller	Modulation	Modal
Feeder	Iteration	Reynolds
Density	Saline	Zircaloy
Estimation	Optimum	Modulus
Hazards	Antenna	Topology
Barriers	Rotor	Zigzag
Coupling	Amplifier	Desalination
Adaptive	Pixel	Outage
Cooling	Deformation	Hydropower
Emission	Configurations	Neutron
Heating	Warehouse	Penstock
Conductivity	Velocities	Neutrons
Loading	Deposition	Analog
Loads	Induced	Doppler
Waves	Magnitude	Fourier
Picker	Electron	Pelton
Machine	Entities	Epoxies
Microhardness	Rotational	Graphene
Substation	Threshold	Hydrophobicity
Transmitter	Residual	Inverter
Agents	Silicon	Boron
Application	Fleet	Isothermal
Steam	Fluid	Alloying
Stiffness	Particle	Anisotropic
Technology	Sodium	Antiscalant
Absorption	Solar	Wavelet
Cycle	Aviation	Macrostructure
Boundary	Axis	Delamination
Capacity	Composites	Detwinned
Components	Cylinder	Linearization
Safety	Degradation	Polarizations
Variation	Cellular	Lognormal
Approximation	Feasibility	Seawater
Tracking	Discrete	Setup
Lean	Distillation	Workflow
Storage	Attenuation	Wimax
Scaling	Compression	Microstrip
Molecular	Fuzzy	Biofouling
Carries	Altitude	Cavitating
Surfaces	Shear	Radiotherapy
Switching	Friction	Ultrafiltrations
Dimensional	Incubators	Throughput
Tool	Kinetic	Orthogonal

Patch	Pendant	Austenite
Experiment	Amplitude	Entropy
Structure	Diameter	Circumferentially
Beams	Voltage	Poiseuille
Permittivity	Humidity	Checksum
Shield	Generators	Instrumented
Quantity	Hollow	Gaussian
Mechanism	Triangular	Anodization
Smoothing	Robot	Tensile
Wiers	Reservoir	Irradiance
Architecture	Welding	Methanol
Reactor	Valves	Dielectric
Magnetic	Corrosion	Annulus
Transformer	Hierarchy	Dialysis
Consumption	Membrane	Demodulator
Tubes	Fossil	Bandwidth
Installation	Hardware	Foulant

Appendix H: YES/NO Receptive Vocabulary Tests

TEST 1

High-Frequency Engineering Technical Vocabulary

1. **Instructions**

This test is used to test your receptive knowledge of technical vocabulary in the engineering domain.

If you know the meaning of the word, tick [√] YES; if you are not sure or don't know the meaning, tick [√] NO.

BE CAREFUL: Not all of the words in the list are 'real' English words; some of them do not exist.

If you tick a word as [√] YES which does not exist, we will deduct one point! The main aim of this test is to measure the breadth of your vocabulary in the engineering domain.

Here is one example:

TRAVERSE [√] YES [ ] NO

LOBE [ ] YES [√] NO

Tunne [ ] YES [√] NO

You know the meaning of the word "traverse" – in this case, you tick YES.

You don't know the word "lobe" – in this case, you tick NO.

You are not sure about the word "tunne" – in this case, also tick NO.

SN	Words	Answer		SN	Words	Answer	
		Yes	No			Yes	No
1	Drill			51	Equiplent		
2	Converter			52	Cycle		
3	Baus			53	Boundary		
4	Curves			54	Capacity		
5	Compulation			55	Exposere		
6	Spacer			56	Components		
7	Filters			57	Entenseaty		
8	Carbon			58	Safety		
9	Loed			59	Variation		
10	Coolant			60	Approximation		
11	Rays			61	Maintananc		
12	Fibers			62	Tracking		
13	Coolang			63	Lean		
14	Rating			64	Storage		
15	Hartnese			65	Modeal		
16	Controller			66	Scaling		
17	Feeder			67	Molecular		
18	Drilly			68	Oparater		
19	Density			69	Carrier		
20	Estimation			70	Surfaces		
21	Hazards			71	Switching		

22	Locadization			72	Pathe		
23	Barriers			73	dimensional		
24	Coupling			74	Tool		
25	Adaptive			75	Reative		
26	Denseatee			76	Patch		
27	Cooling			77	Experiment		
28	Emission			78	Roat		
29	Heating			79	Structure		
30	Conductivity			80	Beams		
31	Noisey			81	Spoed		
32	Loading			82	Permittivity		
33	Loads			83	Shield		
34	Waves			84	Streom		
35	Picker			85	Quantity		
36	Stourege			86	Mechanism		
37	Machine			87	Tecnolonogy		
38	Microhardness			88	Smoothing		
39	Chaneaol			89	Wires		
40	Substation			90	Airdraft		
41	Cyculear			91	Architecture		
42	Transmitter			92	Reactor		
43	Agents			93	Capacitee		
44	Application			94	Magnetic		
45	Concentrazion			95	Transformer		
46	Steam			96	Comcrate		
47	Stiffness			97	Absorption		
48	Technology			98	Tubes		
49	Elactrocal			99	Installation		
50	Absorption			100	Condactivitee		

TEST 2

Mid-Frequency Engineering Technical Vocabulary

1. **Instructions**

This test is used to test your receptive knowledge of technical vocabulary in the engineering domain.

If you know the meaning of the word, tick [√] YES; if you are not sure or don't know the meaning, tick [√] NO.

BE CAREFUL: Not all of the words in the list are 'real' English words; some of them do not exist

If you tick a word as [√] YES which does not exist, we will deduct one point! The main aim of this test is to measure the breadth of your vocabulary in the engineering domain.

Here is one example:

TRAVERSE [√] YES [ ] NO

LOBE [ ] YES [√] NO

Tunne [ ] YES [√] NO

You know the meaning of the word "traverse" – in this case, you tick YES.

You don't know the word "lobe" – in this case, you tick NO.

You are not sure about the word "tunne" – in this case, also tick NO.

SN	Words	Answer		SN	Words	Answer	
		Yes	No			Yes	No
1	Ceramic			51	Solar		
2	Alumeom			52	Valges		
3	Turbine			53	Aviation		
4	Torque			54	Axis		
5	Compate			55	Composites		
6	Electromagnetic			56	Capslate		
7	Armchair			57	Cylinder		
8	Vacuum			58	Degradation		
9	Disglacement			59	Deformination		
10	Condenser			60	Cellular		
11	Pores			61	Feasibility		
12	Aluminum			62	Discrete		
13	Dynamay			63	Distillation		
14	Analyzer			64	Friction		
15	Modulation			65	Dizcretee		
16	Hollawa			66	Incubators		
17	Iteration			67	Kinetic		
18	Saline			68	Pendant		
19	Optimum			69	Fricton		
20	Hybrade			70	Amplitude		
21	Antenna			71	Attenuation		
22	Rotor			72	Compression		

23	Momentame			73	Sperctreal		
24	Pixel			74	Fuzzy		
25	Deformation			75	Altitude		
26	Rotasionale			76	Shear		
27	Configurations			77	Tourbines		
28	Simeolation			78	Diameter		
29	Amplifier			79	Ambent		
30	Warehouse			80	Hardware		
31	Velocities			81	Voltage		
32	Tarminale			82	Humidity		
33	Deposition			83	Gradirt		
34	Induced			84	Generators		
35	Magnitude			85	Inleate		
36	Cereomic			86	Hollow		
37	Electron			87	Triangular		
38	Entities			88	Methanned		
39	Genarato			89	Robot		
40	Rotational			90	Saleane		
41	Threshold			91	Reservoir		
42	Residual			92	Sunnestrate		
43	Noges			93	Welding		
44	Silicon			94	Valves		
45	Fleet			95	Corrosion		
46	Temporeal			96	Porousion		
47	Fluid			97	Hierarchy		
48	Thermarle			98	Rapiol		
49	Particle			99	Membrane		
50	Sodium			100	Fossil		

TEST 3

Low-Frequency Engineering Technical Vocabulary

1. **Instructions**

This test is used to test your receptive knowledge of technical vocabulary in the engineering domain.

If you know the meaning of the word, tick [√] YES; if you are not sure or don't know the meaning, tick [√] NO.

BE CAREFUL: Not all of the words in the list are 'real' English words; some of them do not exist (Pseudowords).

If you tick a word as [√] YES which does not exist, we will deduct one point! The main aim of this test is to measure the breadth of your vocabulary in the engineering domain.

Here is one example:

TRAVERSE [√] YES [ ] NO

LOBE [ ] YES [√] NO

Tunne [ ] YES [√] NO

You know the meaning of the word "traverse" – in this case, you tick YES.

You don't know the word "lobe" – in this case, you tick NO.

You are not sure about the word "tunne" – in this case, also tick NO.

SN	Words	Answer		SN	Words	Answer	
		Yes	No			Yes	No
1	Microfiltration			51	Macrostructure		
2	Modularized			52	Delamination		
3	Nonadaptation			53	Lameoner		
4	Conpection			54	Detwinned		
5	Convection			55	Linearization		
6	Exponential			56	Lognormal		
7	Irragiation			57	Analooge		
8	Speckle			58	Seawater		
9	Axial			59	Setup		
10	Ergiunomic			60	anodyzation		
11	Ergonomic			61	Workflow		
12	Hydroelectric			62	Wimax		
13	Impedance			63	Microstrip		
14	Modal			64	Distayllate		
15	Imdepance			65	Biofouling		
16	Reynolds			66	Cavitating		
17	Zircaloy			67	Dispeersave		
18	Vatax			68	Radiotherapy		
19	Modulus			69	Polarizations		

20	Topology			70	Ultrafiltration		
21	Zigzag			71	Delaminotio		
22	Matanol			72	Throughput		
23	Stoinochiometric			73	Circumbrentially		
24	Desalination			74	Orthogonal		
25	Outage			75	Austenite		
26	Macropostructure			76	Entropy		
27	Hydropower			77	Circumferentially		
28	Neutron			78	Cavytating		
29	Neutrons			79	Poiseuille		
30	Isolathermal			80	Antiscalant		
31	Penstock			81	Checksum		
32	Analog			82	Workflawn		
33	Doppler			83	Instrumented		
34	Fourier			84	Seowatar		
35	Zipzag			85	Gaussian		
36	Pelton			86	Anodization		
37	Epoxies			87	Tensile		
38	Irreogiance			88	Recargeability		
39	Graphene			89	Irradiance		
40	Hydrophobicity			90	Analooge		
41	Inverrtier			91	Methanol		
42	Inverter			92	Dielectric		
43	Boron			93	Noatron		
44	Instrudemented			94	Annulus		
45	Isothermal			95	Dialysis		
46	Alloying			96	Lameoner		
47	Anisotropic			97	Demodulator		
48	Epoxierties			98	Bandwidth		
49	Epithoermale			99	Ortheogonel		
50	Wavelet			100	Foulant		

Good Luck

Appendix I: 90 Pseudowords Selected for Receptive Knowledge Test

90 Pseudowords selected based on the frequency bands classifications high, mid-, low frequency levels as well as supplementary lists.

Words from High-frequency Band			
S/no	Pseudowords	Origin	Level
1	Baus	Bus	1
2	Compulation	computation	1
3	Coolang	cooling	1
4	Loed	Load	1
5	Hartnese	hardness	1
6	Locadization	localization	1
7	Noisey	Noise	1
8	stourege	Storage	1
9	Chaneaol	Channel	2
10	Cyculear	Circular	2
11	Concentrazion	Concentration	2
12	elactrocal	electrical	2
13	equiplent	equipment	2
14	Exposere	Exposure	2
15	Entenseaty	Intensity	2
16	maintananc	maintenance	2
17	Modeal	model	2
18	oparater	operator	2
19	Pathe	path	2
20	Reative	reactive	2
21	Roat	root	2
22	spoed	speed	2
23	streom	stream	2
24	tecnolonogy	technology	2
25	Airdraft	aircraft	3
26	Capacitee	capacity	3
27	comcrate	concrete	3
28	condactivitee	conductivity	3
29	Denseatee	Density	3
30	Drilly	Drill	3

Words from Mid-frequency Bands			
S/no	Pseudowords	Origin	Level
1	alumeom	aluminum	4
2	compate	compact	4
3	disglacement	displacement	4
4	Dynamay	Dynamic	4
5	Hollaway	Hollow	4
6	Hybrade	Hybrid	4
7	Momentame	momentum	4
8	Rotasionale	Rotational	4
9	Simeolation	Simulation	4
10	tarminale	terminal	4
11	cereomic	ceramic	5
12	genarato	generator	5
13	Noges	Nodes	5
14	Temporeal	Temporal	5
15	Thermarle	Thermal	5

16	Valges	Valves	5
17	Capslate	Capsule	6
18	Deformination	Deformation	6
19	dizcretee	discrete	6
20	Fricton	friction	6
21	Sperctreal	spectral	6
22	tourbines	turbines	6
23	ambent	ambient	7
24	Gradirt	gradient	7
25	Inleate	inlet	7
26	methanned	methane	7
27	Saleane	saline	7
28	sunnestrate	substrate	7
29	Porousion	porous	8
30	Rapiol	radial	8

Words from Low-frequency Plus Supplementary Lists			
S/no	Pseudowords	Origin	Level
1	Conpection	convection	9
2	Irragiation	irradiation	9
3	ergiunomic	ergonomic	10
4	imdepance	impedance	10
5	Vatax	vortex	10
6	Matanol	methanol	11
7	zipzag	zigzag	11
8	ortheogonel	orthogonal	12
9	Lameoner	laminar	16
10	Noatron	neutron	17
11	Analooge	analog	25
12	recargeability	reachability	CN
13	seowatar	seawater	CN
14	workflawn	workflow	CN
15	anodyzation	anodization	EVL
16	antiscalant	antiscalant	EVL
17	cavytating	cavitating	EVL
18	circumbrentially	circumferentially	EVL
19	Delaminotio	delamination	EVL
20	Dispeersave	dispersive	EVL
21	Distayllate	distillate	EVL
22	Epithoermale	epithermal	EVL
23	Epoxierties	epoxies	EVL
24	instrudemented	instrumented	EVL
25	Inverrtier	inverter	EVL
26	Irreogiance	irradiance	EVL
27	isolathermal	isothermal	EVL
28	macropostructure	macrostructure	EVL
29	stoinochiometric	stoichiometric	EVL
30	sulfealte	sulfate	EVL